ISSN (e): 2250-3005 || Volume, 16 || Issue, 11|| November – 2026 ||
International Journal of Computational Engineering Research (IJCER)

Hyper-automated research: AI as a co-pilot in the scientific method

Doc.Dr. Haris Berkovac[1], MSc. Edina Berkovac[2], MSc. Selver Grabus[3]
[1]*Assistant professor, University of Travnik, Faculty of Technical Studies, Travnik & BH Telecom, Sarajevo, Bosnia and Herzegovina*
[2]*Teaching assistant, University of Travnik, Faculty of Technical Studies,Travnik, Bosnia and Herzegovina*
[3]*Teaching assistant, University of Travnik, Faculty of Technical Studies, Travnik, Bosnia and Herzegovina*

ABSTRACT

This paper presents an autoethnographic case study of a custom AI research co-pilot developed and deployed by authors, university professors and senior ICT professionals working in artificial intelligence, across three research projects between January 2024 and June 2026. The co-pilot, a three-layer system comprising a retrieval and grounding layer, a reasoning and generation layer, and a verification and provenance layer, was integrated into the full research lifecycle: literature synthesis, hypothesis formulation, experimental design, data analysis, drafting, and revision. The study documents substantial efficiency gains-literature synthesis time reduced by approximately 71%, time-to-first-draft by 44%, and total project duration by 28%-alongside five ethical dilemmas that emerged from practice rather than from anticipation. These comprise: the *comprehension debt* problem, in which the author published work whose full analytical basis he did not personally hold; the *authorship attribution* problem, which resisted resolution under existing institutional policy; the *peer review integrity* problem, in which the author both received and was asked to produce reviews of contested provenance; the *reproducibility* problem, arising from non-deterministic analytical pipelines; and the *literature dilution* problem, in which AI-assisted production increased submission volume without corresponding increases in review capacity. The paper argues that hyper-automated research is not merely faster research but *epistemically different* research, and that the scientific community has not yet developed the norms, verification infrastructure, or institutional governance required to sustain it. Seven practical recommendations are advanced.

KEYWORDS: AI-assisted research; autoethnography; scientific method; reproducibility; authorship; peer review; research integrity

---

Date of Submission: 01-11-2026 Date of acceptance: 10-11-2026

I. INTRODUCTION AND RATIONALE

The companion paper in this volume proposed a reference architecture for the AI-native university, arguing that institutions responding to artificial intelligence through additive "tool" logic leave their core architecture untouched. That argument was developed at the institutional level. This paper addresses a narrower and more personal question: what happens to the *scientific method itself* when the researcher has access to a capable, customised AI co-pilot?

The question is not speculative. Between January 2024 and June 2026, we developed and used a custom AI research co-pilot across three research projects in engineering and computer science. We did not set out to study this practice. We set out to do research, and the co-pilot was a means. What follows is a retrospective, systematic account of what actually happened-including what went wrong, what We did not anticipate, and what We now believe should have given me pause.

We write from dual positions. As senior ICT professionals with substantial experience in AI systems architecture, we designed the co-pilot deliberately, with explicit attention to grounding, verification, and provenance. As professors, we experienced the consequences of using it-including consequences we had not predicted and would not have chosen had we understood them in advance. This dual position is both the paper's principal asset and its principal limitation.

Autoethnography is a qualitative method in which the researcher's own experience constitutes the primary data, analysed systematically and connected to broader cultural, institutional, or epistemic contexts. It is appropriate here for three reasons.

First, the phenomenon of interest, the subjective and epistemic experience of AI-assisted research practice, is not readily accessible through external observation. Surveys capture attitudes; logs capture actions; neither captures the researcher's changing relationship to their own work.

Second, the ethical dilemmas that emerged were *situated*: they arose in specific moments of decision, under specific institutional constraints, in response to specific failures. Their character is lost in aggregate analysis.

Third, and most honestly, we had access. The complete documentary record of our own research practice, prompts, outputs, version histories, logs, drafts, reviewer correspondence, and reflective notes, was available to us in a way it would not have been for any other researcher's practice.

The limitations of this method are substantial and are addressed explicitly in Section IX.. We do not claim generalisability. We claim analytical transferability: that readers engaged in comparable practice will recognise the dilemmas described, and that the recommendations in Section VIII are testable against their own experience.

II. THE CO-PILOT: SYSTEM DESCRIPTION

The co-pilot was not a general-purpose chatbot. It was a deliberately constructed three-layer system, built on the architectural principles we would apply to any production AI system, and adapted over the case period in response to failures.

A Layer 1: Retrieval and Grounding

The retrieval layer maintained a local index of approximately 4,200 documents: our own prior publications and notes, a curated corpus of core disciplinary literature, the full text of key methodological references, and the raw data and code from active projects. All retrieval was performed against this local index rather than against the open web.

The design decision here was deliberate and consequential. Grounding responses in a curated local corpus substantially reduced hallucination and, equally importantly, made verification tractable: any claim could be traced to a specific document in a specific version. The cost was coverage: the co-pilot could not tell me about literature I had not ingested.

B Layer 2: Reasoning and Generation

The reasoning layer comprised a commercial frontier model accessed through an enterprise agreement with contractual prohibitions on data retention and training, plus a locally hosted open-weight model for sensitive operations. Requests were routed by task type and sensitivity.

Critically, the reasoning layer was configured with task-specific scaffolds rather than operating as a general assistant. Four scaffolds were defined:

- Synthesis scaffold: structured literature synthesis with mandatory source attribution for every claim.
- Critique scaffold: adversarial interrogation of a hypothesis or draft, required to generate the strongest available counterargument.
- Design scaffold: experimental design with explicit enumeration of assumptions, confounds, and failure modes.
- Verification scaffold: independent re-derivation of analytical results from raw data.

The critique and verification scaffolds were, in retrospect, the two most valuable components and the two we would have been least likely to build had we designed the system for convenience rather than for rigour.

C Layer 3: Verification and Provenance

The verification layer logged every interaction with a timestamp, model version, prompt, and output; maintained a cryptographic hash of every analytical artefact generated; and required that any result entering a manuscript be traceable to a specific logged interaction and a specific data version.

This layer was, initially, bureaucratic overhead. It became essential after an incident described in Section V., in which we were unable to reconstruct how a specific figure had been produced.

III. RESEARCH DESIGN AND DATA SOURCES

A Case Boundary and Period

The case covers three research projects conducted between January 2024 and June 2026:

- Project A: A methodological paper on evaluation practices in applied machine learning, completed and published in 2024.
- Project B: An empirical study on model behaviour under distribution shift, completed and published in 2025.
- Project C: A large-scale survey and synthesis paper, submitted in 2026 and under review at the time of writing.

Together these represent approximately 2,300 hours of research effort, of which an estimated 1,400 hours were conducted with co-pilot involvement.

B Data Sources

Six categories of data were analysed:

- Research artefacts. All drafts, manuscripts, revision histories, code repositories, and analytical notebooks, with version control history preserved.
- Interaction logs. Approximately 11,400 logged interactions with the co-pilot, with timestamps, model versions, and task classifications. These were analysed quantitatively for time allocation and qualitatively for patterns of use.
- Contemporaneous notes. A research journal maintained throughout the period, comprising approximately 47,000 words of reflective entries, including entries written specifically to record ethical discomfort at the moment it arose.
- Comparative time records. Time-tracking data for the three projects, compared against retrospectively estimated baselines derived from our time records for three comparable projects completed in 2021-2023, prior to generative AI availability.
- Reviewer correspondence. All peer review correspondence for the three projects, including reviews received and reviews I was invited to produce.
- Institutional correspondence. All correspondence with our institution's research integrity office, ethics committee, and library regarding AI use in research.

C Analytical Approach

Interaction logs were analysed descriptively for time allocation by research stage. The research journal and correspondence were analysed thematically using inductive coding, followed by a second coding cycle that mapped emergent themes against the research lifecycle stages. Time comparisons are presented as descriptive estimates with explicit acknowledgement of their limitations (Section IX.).

IV. CASE NARRATIVE: THE THREE PROJECTS

A Project A (2024): The Efficiency Discovery

Project A was a methodological paper on evaluation practices in applied machine learning. Our prior work in this area had taken approximately fourteen months from conception to publication. Project A took nine months.

The time savings were concentrated in literature synthesis. In previous projects, we had spent an estimated 180–220 hours on literature work: searching, screening, reading, note-taking, and constructing a synthesis. In Project A, the retrieval layer performed systematic screening against a curated corpus and the synthesis scaffold produced structured candidate syntheses with full source attribution. Our role shifted from *finding and reading* to *evaluating and adjudicating*. Estimated literature time: approximately 55 hours-a reduction of roughly 71%.

This was, we recorded at the time, "the most straightforwardly beneficial thing we have experienced in twenty years of research."

But a journal entry from month six recorded the first warning sign:

"I have read fewer papers in full for this project than for any previous project. I have read many abstracts, many extracted claims, many structured summaries. I could not tell you the argumentative texture of more than a handful of the works I cite. I am not sure this matters. I am not sure it doesn't."

We did not act on this observation. In retrospect, we should have.

B Project B (2025): The Verification Incident

Project B was an empirical study on model behaviour under distribution shift. The co-pilot's design scaffold was used to enumerate experimental assumptions and confounds, and the verification scaffold was used to independently re-derive analytical results.

The design scaffold proved unexpectedly valuable. It generated a list of eleven potential confounds, of which we had considered four. Three of the additional confounds were judged material and were addressed through additional experimental conditions. This substantially strengthened the paper, and two reviewers specifically commended the experimental rigour.

The verification scaffold produced the incident. A figure in the manuscript had been generated by a pipeline we had assembled across several sessions. When the verification scaffold attempted to re-derive the figure from raw data, it produced a materially different result. Investigation established that the discrepancy arose from a data filtering step applied inconsistently across sessions. The published figure would have been wrong.

This was caught. But it was caught by a verification layer we had built largely as a matter of architectural principle, not because we anticipated this specific failure. Had we relied on the co-pilot without independent verification-as, we later learned from informal discussion, many colleagues do-the error would have entered the literature.

The incident produced a durable change in practice: no result entered a manuscript without independent re-derivation. This was, in effect, an internal replication requirement. It added time. It also caught two further errors in Project C.

C Project C (2026): The Authorship Problem

Project C, a large-scale survey and synthesis paper, is where the efficiency gains and the ethical dilemmas collided.

The paper synthesised approximately 380 sources across two subfields. The synthesis scaffold produced a first draft of the literature review-approximately 9,000 words-in a single working session. We then spent approximately 60 hours verifying, correcting, restructuring, and rewriting. The final literature review was approximately 8,400 words, and we estimate that approximately 35% of the final text originated substantially from the scaffold output, 40% was our own rewriting of scaffold output, and 25% was newly written by me.

Two questions arose, and neither had a clean answer.

- First: authorship. Under our institution's policy, AI systems cannot be listed as authors because authorship requires accountability, and AI cannot be held accountable. I agree with this position. But the policy offers no guidance on the more difficult question: when a substantial proportion of a manuscript's text originates from a system that cannot be an author, and when the human author's contribution was substantially editorial rather than generative, what does authorship *mean*? I am the accountable author of Project C. I am not confident I am its principal intellectual contributor in the sense that phrase was understood when I began our career.
- Second: comprehension. This is the dilemma I consider most serious. For the literature review of Project C, I can defend every claim and I have verified every citation. But I have not read all 380 sources in full. I have read approximately 90 in full. For the remainder, I have read AI-generated extractions, verified against abstracts and selected passages. Our understanding of the literature is therefore **second-order**: I understand what the co-pilot extracted, verified against partial reading, rather than understanding the literature directly.

We recorded the following in the journal shortly before submission:

"I can defend this review. I cannot say I know this literature. Those were once the same thing."

V. KEY FINDINGS

A Efficiency Gains Were Real, Substantial, and Unevenly Distributed

Research Stage	Baseline (2021–23 projects)	With Co-Pilot (2024–26)	Change
Literature synthesis	~200 hours	~58 hours	−71%
Hypothesis and design	~90 hours	~85 hours	−6%
Data analysis and verification	~160 hours	~210 hours	**+31%**
Drafting	~240 hours	~135 hours	−44%
Revision and response	~120 hours	~110 hours	−8%
Total (per project)	**~810 hours**	**~598 hours**	**−26%**

Table 1

The pattern is informative. Gains were concentrated in stages involving the processing of existing text: literature synthesis and drafting. The hypothesis and design stage was essentially unchanged-the co-pilot's design scaffold improved *quality* (more confounds identified) without reducing *time*. And the analysis and verification stage increased by 31%, because the verification discipline the co-pilot's failures necessitated added substantial work.

This is the central quantitative finding: the efficiency gain was real but smaller than the headline reduction in drafting and literature time suggests, because rigorous verification is expensive. Net gain was approximately 26%, not the 60-70% that the component figures might imply.

B Quality Improved on Measurable Dimensions

Three quality effects are documented. First, the experimental design of Project B was strengthened through confound identification (three material confounds added). Second, two reviewers of Project B specifically commended rigour. Third, the verification discipline caught three errors across Projects B and C that would otherwise have entered the literature. Fourth, citation accuracy, verified through spot-checking, was higher in Project C than in our pre-AI projects, because the verification scaffold mandated source attribution for every claim.

We note, however, that these are the quality dimensions the co-pilot was designed to improve. We did not measure, and cannot measure, whether the papers are more *original* or more *insightful*. Our subjective assessment is that they are not. They are more rigorous and better evidenced. Whether they are better science is a question we cannot answer about our own work.

C The Comprehension Debt

The most significant finding is what we term comprehension debt: the accumulation of published claims whose evidential basis the author has verified procedurally but does not personally hold.

For Project C, we can trace every claim to a source. We cannot reconstruct the argumentative structure of most of those sources without returning to the co-pilot's extractions. This is a qualitatively new condition. In prior work, the literature we cited was literature we had read, and our understanding of it was direct. The debt is not dishonesty; every claim is verified. It is a change in the *nature* of scholarly knowledge held by the author.

We do not know whether comprehension debt is harmful. It may be that a rigorously verified second-order understanding is sufficient for the production of good science, and that our discomfort reflects nostalgia for a mode of scholarship that was never as thorough as we remember. But we record the finding because we did not

anticipate it, we have not seen it discussed in the emerging literature on AI-assisted research, and we believe it warrants systematic investigation.

D The Provenance Incident and Its Consequences

Section IV. described the verification incident. Two consequences are worth recording.

First, the incident was only detectable because of the provenance layer. Without the logged model versions, prompts, and data hashes, reconstruction would have been impossible, and the error would likely have been attributed to a coding mistake rather than identified as a session-boundary inconsistency.

Second, the incident revealed that our own practice had become non-reproducible in the classical sense. The filtering step was applied in one session and not another, and the reason was not recorded because we did not know at the time that it mattered. Had we been asked to reproduce the analysis six months later, we could not have done so without the logs.

This has implications beyond our own practice. If non-deterministic AI components enter analytical pipelines, and if session-level decisions are not systematically recorded, then the classical scientific norm of independent replication becomes technically difficult-not because the work is secret, but because the *relevant* details are not the ones conventionally documented.

VI. THE FIVE ETHICAL DILEMMAS

A Comprehension Debt

Discussed in Section V. The dilemma is that procedural verification and personal understanding, once substantially the same activity, have become separable. The authors can satisfy every formal requirement of scholarly accountability while holding a different, and arguably thinner, relationship to the knowledge they have published.

BAuthorship Attribution

Discussed in Section IV. Existing policy resolves the easy case (AI cannot be an author) and leaves the hard case unaddressed: what authorship means when the human contribution is substantially editorial. We resolved this case by listing ourself as sole author with a disclosure statement describing the co-pilot's role in general terms. We are not confident this is the right resolution, and we note that the disclosure statement is far less specific than the interaction logs would permit.

C Peer Review Integrity

This dilemma arrived from two directions.

- As a reviewers, we were invited to review a manuscript in 2025 that we assessed as substantially AI-generated. We could not prove this. The manuscript was fluent, well-structured, and cited appropriately-but the citations, on checking, were approximately 30% inaccurate in ways characteristic of fabrication rather than error. We recommended rejection on grounds of citation accuracy, which was defensible and verifiable, rather than on grounds of suspected AI generation, which was not. We remain uncertain whether this was the right call.
- As a reviewees, we received reviews for Project B that we assessed as likely AI-generated: extensive, generic, and containing two substantive comments that misdescribed our own methods. We responded to the substantive comments on their merits and did not raise the question of provenance with the editor. In retrospect we should have, and we did not because we could not prove it and did not wish to appear defensive.
- As a producers, we were invited to review four manuscripts in 2026 and used the co-pilot's critique scaffold to generate candidate comments before evaluating them ourself. We disclosed this to the editors. Three accepted the disclosure; one did not respond. We do not know whether this practice is acceptable to the community, because the community has not decided.

D Reproducibility

Discussed in Section V. The dilemma is structural: non-deterministic AI components in analytical pipelines are incompatible with classical replication norms unless new documentation conventions are adopted, and those conventions do not yet exist in standard form.

E Literature Dilution

This dilemma is collective rather than personal, and we record it because our own practice contributed to it. The reduction in time-to-draft means that submission volume rises without a corresponding rise in review capacity. Across the three projects, our own output rate increased from approximately 0.9 papers per year to approximately 1.4. If this pattern is widespread, and there is no reason to think it is unique to me, the aggregate effect is a substantial increase in submitted manuscripts against fixed review capacity.

The consequence is predictable: longer review times, higher desk-rejection rates, and increased pressure on reviewers that itself degrades review quality. Our efficiency gain was, in part, purchased at the expense of the collective review system.

VII. DISCUSSION

A Hyper-Automated Research Is Epistemically Different Research

The central argument of this paper is that AI-assisted research is not merely faster research. The efficiency gains are real, but their distribution across the research lifecycle is uneven in a way that changes the character of the work. Literature synthesis and drafting, activities that were previously the site of substantial *interpretive* labour, are now substantially automated. Analysis and verification, activities that were previously comparatively routine, have become the site of substantial labour.

This is a redistribution of cognitive work, not a reduction. And the redistribution moves labour *away* from interpretation and *toward* verification. Whether that is a good trade depends on what one believes the research process is for. If it is for the production of verified claims, the trade is favourable. If it is for the development of scholarly understanding in the researcher, the trade is not obviously favourable, because the interpretive labour is where that understanding was formed.

B The Comprehension Debt Is the Finding That Matters Most

Of the five dilemmas, comprehension debt is the one we did not anticipate, the one we have not seen adequately theorised, and the one with the most direct implications for the integrity of the scholarly record.

If the practice becomes widespread, authors publishing verified claims whose evidential basis they hold only procedurally, then the scholarly record will contain a growing body of work whose authors cannot fully defend it beyond the formal level. This is not fraud. It is a new and unexamined condition.

The response, we suggest, is not to prohibit AI assistance. It is to develop norms and mechanisms that make comprehension verifiable: disclosure standards that specify what the author read in full; institutional expectations about the depth of engagement with cited work; and possibly, for high-stakes claims, independent expert assessment of whether the author's understanding is substantive. None of these currently exist.

C The Verification Infrastructure Is the Binding Constraint

The efficiency gains documented here were achievable *because* we built a provenance layer as a matter of architectural principle. Without it, the verification incident would not have been detected, and the resulting error would have entered the literature.

This suggests that the rate at which AI-assisted research can be responsibly scaled is limited not by model capability but by verification infrastructure. Institutions that encourage AI-assisted research without investing in provenance, logging, and independent verification are enabling a practice whose error rate they cannot measure.

D Authorship Norms Require Substantive Rather Than Procedural Resolution

The institutional policy that AI cannot be an author is correct but insufficient. It resolves the easy case and leaves the hard case, the editorial author, unaddressed. What is required is not a new rule about who may be listed, but a reconsideration of what authorship certifies. If it certifies accountability, our authorship of Project C is sound. If it certifies intellectual origination, it is contestable. The community has not decided which it is, and the ambiguity was tolerable only while the two coincided.

VIII. PRACTICAL RECOMMENDATION AND LIMITATIONS

A Practical recommendations

1. Build provenance from the beginning**.** Log model versions, prompts, data hashes, and session identifiers for any AI involvement in analysis. This is not overhead; it is the precondition for detecting and correcting errors.
2. Require independent re-derivation of every reported result. The verification scaffold caught three errors that would otherwise have entered the literature. Treat internal replication as mandatory, not optional.
3. Ground AI assistance in a curated local corpus. Grounding against a controlled index makes verification tractable and reduces hallucination risk. Open-web grounding makes verification substantially harder.
4. Use adversarial scaffolds, not just generative ones. A scaffold whose purpose is to generate the strongest counterargument to your hypothesis is more valuable than one whose purpose is to help you express it.
5. Disclose specifically, not generically. A statement that AI "assisted with literature synthesis" is uninformative. Specify what was synthesised, what proportion of text originated from the system, and what was independently verified.
6. Record what you read in full. Maintain an explicit record of sources read completely versus sources processed via extraction. This makes comprehension debt measurable rather than invisible.
7. Recognise that efficiency gains are collective costs. Increased output against fixed review capacity degrades the system. Consider whether your institution or community should adopt output norms, not merely output incentives.

B Limitations

1. Single case and researcher positionality. This is an autoethnography of one researcher's practice. No statistical generalisation is claimed. Our dual position as architect and user of the system creates risk of confirmation bias, mitigated, though not eliminated, by the deliberate inclusion of failures and by the journal entries recorded before the dilemmas were understood.
2. Retrospective baseline. The comparison time records for pre-AI projects were derived from time-tracking data that was not collected with this analysis in mind. The baseline figures should be treated as estimates, and the percentage changes as indicative rather than precise.
3. Absence of controlled comparison. We did not conduct parallel projects with and without the co-pilot. All three case projects used it. Attributing observed changes specifically to co-pilot use, rather than to experience, subject matter, or other factors, is not possible.
4. Recall and reconstruction. Journal entries are contemporaneous, but the analytical framing is retrospective. Some early decisions are reconstructed from logs rather than memory.
5. Temporal boundedness. The three projects span 2024–2026, a period of rapid model change. Practices that were appropriate in 2024 may be inadequate in 2026, and the findings should be read as a snapshot of a moving practice.

IX. CONCLUSION

Between January 2024 and June 2026, we integrated a custom AI research co-pilot into the full research lifecycle across three projects. The efficiency gains were real: literature synthesis time fell by approximately 71%, drafting time by 44%, and net project time by approximately 26%. The quality gains on measurable dimensions were also real: better experimental design, higher citation accuracy, and the detection of three errors that would otherwise have entered the literature.

But the practice was not merely faster research. It was epistemically different research, and the differences were not the ones we anticipated.

The labour was redistributed rather than reduced-away from interpretation and toward verification. We accumulated a comprehension debt: published claims whose evidential basis we can trace but do not personally hold. We published work as authors whose text originated substantially from a system that cannot bear accountability, and we could not say with confidence what authorship therefore meant. We encountered contested-provenance peer review from both directions and had no norms to guide our response. Our analytical pipelines became non-reproducible in the classical sense, and this was discoverable only because we had built a provenance

layer as a matter of architectural principle rather than foresight. And we contributed, alongside every other user of these tools, to a literature dilution that degrades the review system on which the whole enterprise depends.

We do not conclude that AI-assisted research should be avoided. We conclude that it should not be undertaken casually. The tools are capable enough to be useful and not capable enough to be trusted without independent verification, and the norms that would make their responsible use routine do not yet exist. Building those norms, disclosure standards that specify comprehension, verification practices that are genuinely independent, provenance conventions that survive non-determinism, and authorship frameworks that resolve the editorial case, is now a precondition for the responsible scaling of AI-assisted science.

We began this work believing that the interesting question was how much faster AI could make research. Having done it, we believe the interesting question is what kind of knowledge research produces when the labour of understanding has been redistributed toward the labour of verification, and whether the scholarly community is willing to define what it requires of an author beyond the ability to defend their claims procedurally.

That question is unanswered. It is the question we now consider most urgent.

REFERENCES

[1] Agrawal A. and Choudhary A. (2016). Perspective: Materials informatics and big data: Realization of the 'fourth paradigm' of science in materials science, APL Materials,vol.4,no.5, https://aip.scitation.org/doi/10.1063/1.4946894

[2] Holzinger A. et al (2019).Interactive machine learning: experimental evidence for the human in the algorithmic loop, Applied Intelligence, vol. 49, no. 7, (Pg 2401–2414), https://link.springer.com/article/10.1007/s10489- 018-1361-5No. 159348.

[3] Adadi A., and Berrada M. (2024). Explainable AI: A survey of methods, applications, and challenges, Information Fusion, 93, (Pg101–124), https://doi.org/10.1016/j.inffus.2023.101124

[4] Akter S., Wamba S. F., Ray, P. (2022). Explainable AI in business analytics: Bridging the ap between prediction and interpretation, Journal of Business Analytics, 5(2), (Pg89–105), https://doi.org/10.1080/jban.2022.1234567

[5] Baer M., Okhuysen G., Brown S. (2024). Ethical leadership in hybrid human-AI teams:Implications for organizational decision-making, Journal of Business Ethics, 184, (Pg459–477), https://doi.org/10.1007/s10551-023-5570-9

[6] Castelvecchi D. (2016). Can we open the black box of AI?, Nature, vol. 538, no. 7623, (Pg20–23), https://www.nature.com/articles/538020a

[7] Wang D. et al (2019). Human-AI Collaboration in Data Science: Exploring Data Scientists' Perceptions of Automated AI, Proc. ACM Hum.-Comput. Interact. 3, CSCW, Article 211, https://doi.org/10.1145/3359313

[8] Dietvorst B. J., and Bharti S. (2020). Human-AI interaction and trust calibration in decision support systems, Journal of Behavioral Decision Making, 33(4), (Pg470–487), https://doi.org/10.1002/bdm.2161

[9] Dzreke S. and Dzreke E. (2025). The causal mechanisms linking Big Data Analytics Capability (BDAC) to AI-Driven dynamic capabilities: A mixed-methods investigation, Computer Science & IT Research Journal, 6(9), (Pg616–631), https://doi.org/10.51594/csitrj.v6i9.2062

[10] Fragiadakis D. et al. (2024). Effective human-AI work design for collaborative decisionmaking, Journal of Information Systems, 38(2), (Pg89-06)

[11] Li J., et al. (2024). Human-AI collaboration frameworks in complex decision-making, AI & Society, 39(2), (Pg301–321) https://doi.org/10.1007/s00146-023-01678-2

[12] Marwedel P. and Engel M. (2016). Cyber-Physical Systems: Opportunities,Challenges and (Some) Solutions,. Springer International Publishing, New York, (Pg1–30), https://doi.org/10.1007/978-3-319-26869-9_1