Stout in Business and Management (SBM)
ISSN: 2979-2827
DOI: https://doi.org/10.61424/rj1dhn51
Journal Homepage: https://stoutjournals.org/index.php/SBM

The Augmentation Trap in Analytics Teams: Short-Run AI Productivity, Skill Erosion, and Decision Resilience

Md. Nabidul Haque
Department of Marketing, University of Dhaka, Dhaka, Bangladesh
Corresponding Author: Md. Nabidul Haque, **E-mail**: haquenabid01@gmail.com

ARTICLE INFORMATION

Submitted: March 15, 2025
Accepted: June 20, 2025
Published: November 26, 2025
Volume: 1
Issue: 1

KEYWORDS

Artificial Intelligence, Human-AI Collaboration, Skill Erosion, Automation Bias, Decision Resilience.

ABSTRACT

AI can make analytics teams faster, but it can also be faster without human ability. A concept of augmentation trap is developed and tested in this study. The fallacy is that continuously using AI will enhance existing productions but will also decrease the ability to do analysis without AI and make people overly reliant on the AI system. The paper establishes the connections between human-AI complementarity, experiential learning, cognitive offloading, automation bias and organizational resilience. A twelve week-long quasi-experiment is suggested, which is a mixed-methods approach, and with 48 analytics teams and 192 analysts. Teams are divided into groups that have unrestricted AI access, periodic unaided practice and verification with calibrated AI access, delayed or withheld AI access. The design features repeated assessments without AI, an unannounced model-outage exercise, and a task that has a plausible but erroneous AI recommendation. Multilevel growth models test the changes in productivity and unaided skill. Cognitive offloading and team-level models of decision resilience are the focus of mediation tests. The synthetic analysis is only included as a reporting template. It shows an example of a trend where unrestricted access to AI increases throughput, but leads to poorer unaided skill, outage performance and error detection. When calibrated, it will yield smaller gains in speed and higher long run capability. The study adds a time-based perspective to the discussion of AI augmentation, and it reveals why companies should assess performance and capability with and without AI, as well as with incorrect AI.

1. Introduction

AI is revolutionizing the way organizations process information, forecast, create reports and make decisions. Generative AI and machine-learning tools are now being employed by analytics teams to generate code, scrub data, collate evidence, identify patterns, and formulate management recommendations. These tools can minimize the mundane tasks and save on the time required to accomplish a complicated task. Field and experimental evidence also demonstrates improvements in speed and quality of many types of knowledge work. In their study, Noy and Zhang (2023) discovered significant improvements in productivity for professional writing tasks. Brynjolfsson, Li, and Raymond (2025) discovered that the AI assistant boosted customer-support agents' productivity, especially for those who were less experienced. Additionally, Dell'Acqua et al. (2026) discovered that AI enhanced the performance within its scope of action but declined outside of it.

The results confirm the idea that AI is a tool to complement, not replace, human labor. In an augmentation model, AI provides 4 key elements: scale, search, pattern recognition and fast generation. Human workers to provide context, judgment, values and accountability. The goal is to get a whole greater than the sum of its parts. However, complementarity is not a given. Vaccaro, Almaatouq, and Malone (2024) conducted a meta-analysis of the literature and found that human-AI combinations tended to outperform the human, but they did not perform better than the stronger of the two components, on average. In particular, decision tasks were more difficult. From this perspective, differences in information and capability are thus critical to create effective complementarities, and how tasks and roles are structured are also important in this regard, as argued by Hemmer et al. (2025).

Copyright: © 2025 The Author(s). This open access article is distributed under the terms and conditions of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC- BY-NC-SA) License (https://creativecommons.org/licenses/by-nc-sa/4.0/), published by Stout Journals, London, United Kingdom

Secondly, the time is a problem. Most studies on AI measure output with the AI available. They wonder if users are able to finish faster, do more work, or get better results while assisted in their performance. These measures are crucial but do not provide information about workers' capacity development. An AI system can create great work, but someone might be better able to do all of these without the AI system, and they may not be as good with it. This is a key difference in analytics jobs as need more than outputs to make good decisions. Analysts need to be aware of bad data, question weak assumptions, be aware of model limitations, and act when automated support doesn't help.

This issue has been known for decades in the realm of research on automation. Bainbridge (1983) pointed out that automation may take people out of the routine and leave them in charge during rare and challenging failures. Parasuraman and Riley (1997) discussed misuse, disuse and abuse of automation. Lee and See (2004) demonstrated that it is not enough to rely on trust; it should be 'tuned'. Additional, more recent, research in information-systems also reveals the potential for cycles of skill erosion that can result when users rely on information systems, become complacent, and have poor review practices. For instance, Rinta-Kahila et al. (2023) determined that cognitive automation can decrease mindful task performance and mask decreasing expertise from both workers and supervisors.

The risks might be accentuated by the use of Generative AI for tasks that involve visible aspects of reasoning, not just routine physical or numerical tasks. It may recommend the problem frame, pick a method, write the code, interpret the outcomes, and come up with a refined recommendation. This kind of assistance can come in handy. It can also help ease the struggle, correction of errors and reflection where in which expertise is formed. Recent experiments on the learning of skills indicate that people who offload their learning of unfamiliar skills entirely to AI might be able to increase their output without increasing their understanding (Shen & Tamkin, 2026). Caosun and Aral (2026) refer to a dynamic situation in which immediate skill improvement from AI outweighs the apparent short-term costs, but the skill gradually diminishes as it is partly lost over time, as an augmentation trap. Their model is a good first step, but there is not a lot of evidence at the team level.

In the current study, this issue is to be taken to the teams that have analytics and to operational decisions. It asks: When is AI augmentation an enhancer of analytics capability and when is it an eroder of skills and reliance on operations? There are three questions which follow that relate to this. First, is there any long-term trend in immediate productivity and unaided analytical performance as a result of repeated use of AI? Second, do cognitive offloading account for any loss of unaided skill? Third, how does the teams fare when AI is not present or gives a possible but incorrect recommendation?

The proposed study is a quasi-experimental study with a longitudinal design and mixed-methods that will last for 12 weeks. They get unlimited access to AI, periodic unaided practice and verification or they get delayed access to AI. The design takes into account normal work performance (without AI) and periodic performance (without AI). It also has an Unannounced Outage exercise and an Incorrect-recommendation task. This enables the study to differentiate between the gains in production and capability development. It also evaluates decision resilience, which is used here to describe the capacity to make decisions of an acceptable quality in the absence of AI, in the event of a delay, degradation, or incorrect answer from AI.

The paper makes four contributions. First, it brings temporal aspects to the table of human-AI complementarity. AI could be used to augment current output and replace learning process that underpins future output. Secondly, it distinguishes between performance that is assisted and performance that is merely the result of ability. This difference is a key reason why a good team with AI can end up being fragile. Thirdly, it puts cognitive offloading and experiential learning within an organizational framework of using artificial intelligence. Fourthly, it considers system failure as test condition and not a rare technical occurrence. The key point of the main practical claim is straightforward: Organizations should not assess AI based on what people do with AI. They should also assess the potential of the population to remain functional without it.

2. Literature Review and Theoretical Development

2.1 AI Augmentation and Short-Run Productivity

AI augmentation is where the capabilities of humans and machines are combined in a single process. It's a nice logic in analytics. AI can sift through expansive data sets, write code, identify patterns, and make quick drafts. Analysts are able to establish objectives, weigh context, consider trade-offs, and be responsible for the final decision. Jarrahi (2018) termed this relationship as human-AI symbiosis in decision making process in the organization. Dellermann et al. (2019) applied the broader concept of hybrid intelligence that is the fusion of human and machine capabilities to accomplish tasks in which neither is particularly adept. There is now unequivocal evidence of short run benefits. Noy and Zhang (2023) proved that on writing tasks completion time was decreased and quality increased due to the use of generative AI. Brynjolfsson et. al. (2025) observed an average of 15% more issues being resolved per hour by 5,172 support agents. Recent and less-skilled workers benefited more, indicating the potential of AI to disseminate beneficial practices. Dell'Acqua et al. (2026) have identified that consultants using AI are more likely to finish more tasks, work faster, and create higher quality work on tasks deemed to be in the AI frontier. In a second field

experiment, AI was found to be similar to a team without AI in terms of product innovation, and AI was used to integrate technical and commercial perspectives (Dell'Acqua, Ayoubi, et al., 2026).

These studies demonstrate useful, but also demonstrate limitations. While some studies have shown that AI can enhance correctness on a given task, others have reported that it can lead to lower correctness on a task that is not in its frontier, as found by Dell'Acqua et al. (2026). Vaccaro et al. (2024) concluded that the difference in performance between human and AI was dependent on the type of task and the relative strength of the human and the AI. According to Zercher et al. (2025), the effectiveness of AI was found to be higher when the new information it provided helped to mitigate information gaps among team members, but faded when the information gap was created by the introduction of new information related to AI. Therefore, access to AI is not sufficient to make for an effective augmentation. Tasks need to be allocated properly, and monitored and learnt in the work system.

This paper distinguishes between two types of complementarities. Production complementarity occurs when the production of the human-AI unit is increased by the use of AI. Capability complementarity also occurs when AI supports the further development of the knowledge and judgment that can be applied subsequently, even in the absence of AI. Production complementarity can happen in a rapid manner. It takes time to develop capacity complementarity, and active engagement, feedback, and retained practice. The two forms can travel along together or opposite directions.

2.2 Experiential Learning, Cognitive Offloading, And Skill Erosion

Analytical skills are acquired with experience. People learn when they formulate problems, try out procedures, get feedback, fix their errors, and transfer what they've learned to other procedures. Experiential learning doesn't just happen in formal learning. It has been integrated into the program of repetitive work. Organizational learning is change in knowledge which occurs after experience, as described by Argote and Miron-Spektor (2011). March (1991) is also a reminder of the need to exploit the known methods but also to explore other methods that will enable firms to improve their future capabilities.

AI can serve as an aid to learning by explaining, asking questions, giving feedback or introducing workers to improved techniques. Shao et al. (2025) reported that the knowledge gain and task performance may be enhanced by daily use of augmentation-based AI. Information overload and recovery costs, however, were revealed in the same study. The impact of AI is thus contingent on its usage. When a worker takes the idea to the test with AI, the worker is still cognitively engaged. If a worker asks AI to solve a problem, perform analysis, and compose a conclusion, he might do the task without practicing much of the skill.

Cognitive offloading is the process of delegating memory, calculation, search or reasoning to an external support. Offloading doesn't necessarily do more harm than good. Low-value effort can be removed from the human process by the use of calculators, software and checklists so that people can concentrate on the higher level of judgment. The risk comes into play when the system replaces the cognitive actions that workers must grasp, monitor or retrieve. AI supportive offloading is where AI takes away the toil and repetition of the task while the analyst still frames, checks, and explains. Delegative offloading core reasoning is delegated to the analyst who will mainly accept or edit the output.

Skill erosion is a reduction in the skills that are required for relevant work without the aid. It can impact on method choice, error diagnosis, error interpretation, coding or transfer to a different context. Rinta-Kahila et al. (2023) show that skill erosion can form a vicious circle. If people rely more on their computer, their mindful performance will decrease; if their mindful performance is less, their skill will decrease; and if their skill decreases, their need for automation will increase. Recent experiments by Shen and Tamkin (2026) offer experimental evidence of the adverse effects of heavy delegation to AI on conceptual understanding and debugging skill. Their research further indicates that learning can be maintained with the utilization of AI, despite the presence of engaged interaction patterns.

The essence of skill erosion is that it is a dynamic process. It's not something that can be seen through one assisted task. This study should make comparisons between baseline skill and repeated unaided assessments. Should also use practice, fatigue and task difficulty as control variables. In order to get a valid test, then the interaction between treatment and time should be evaluated in unaided performance.

2.3 Automation Bias, Trust and Dependence on Automation

Automation bias is when individuals take automated recommendations at face value or do not look for information that contradicts them. Skitka, Mosier, and Burdick (1999) demonstrated that automated decision aids have the potential to commit omission and commission errors. It has also been shown that the effectiveness and characteristics of automation influence human performance (Alla, 2025). Trust forms such a part of this process. Low trust means they may not take advantage of helpful

support and high trust can result in taking on poor output. Trust and capabilities in the system are both necessary for appropriate reliance (Mehrotra et al., 2024).

Calibration is a challenge due to the fluency of generative AI system output, even when they are incorrect. A well-worded explanation might lessen the need for confirmation by the user. Explanations do not necessarily lead to better joint function. Bansal et al. (2021) discovered that explanations could be used to help the understanding of a model but may not lead to the complementary performance of teams. Logg, Minson, and Moore (2019) demonstrate that people can become algorithm appreciative, and Dietvorst, Simmons, and Massey (2018) demonstrate that they can become algorithm averse by observing mistakes. It is possible that both reactions may be poorly calibrated.

This study is meant to differentiate between the use of AI and AI dependence. Heavy use does NOT equal dependence. A team can continuously leverage AI without compromising their ability to stay resilient when they have key knowledge and skills to review the outputs of AI tools and maintain manual tasks. When a large loss in performance occurs on the removal or failure of the AI, then there exists a dependence. That loss is a measure of the size. It can manifest itself in terms of less accuracy, slower completion, weaker recovery, inadequate confidence calibration or not detecting that a bad recommendation has been made.

2.4 Team Learning and Decision Resilience

Analytics work is best done by teams as no one knows all of the relevant facts. The sharing of knowledge, communication and finding the right expertise within the team are key factors in the capability of the team. Lewis (2003) defines transactive memory as a 'shared system for knowing who knows what. AI can bolster this system with knowledge organization and providing a shared point of support. It can also weaken it when all members turn their questions to the same model and they cease to learn from colleagues.

The latest research takes a different approach to AI, considering it not just a tool, but a teammate. AI teaming impacts on coordination, trust and understanding, and the value of AI teaming depends on team design and knowledge configuration of the AI (Schmutz et al., 2024; Seeber et al., 2020; Zercher et al., 2025). Less is known about performance following withdrawal of AI or failure of the AI.

Organizational resilience is the ability of an organization to predict, withstand and respond to disruptions (Duchek, 2020). HR systems can also build up a reservoir of flexible skills and behaviors that foster resilience (Li & Lin, 2024). In an AI-powered analytics team, decision resilience goes beyond just having a backup server. It demands individuals who can think through things, identify untrustworthy output, operate within a coordination framework in the event of stress and recover performance. Outage and error are considered as stress tests of the joint system for the present study.

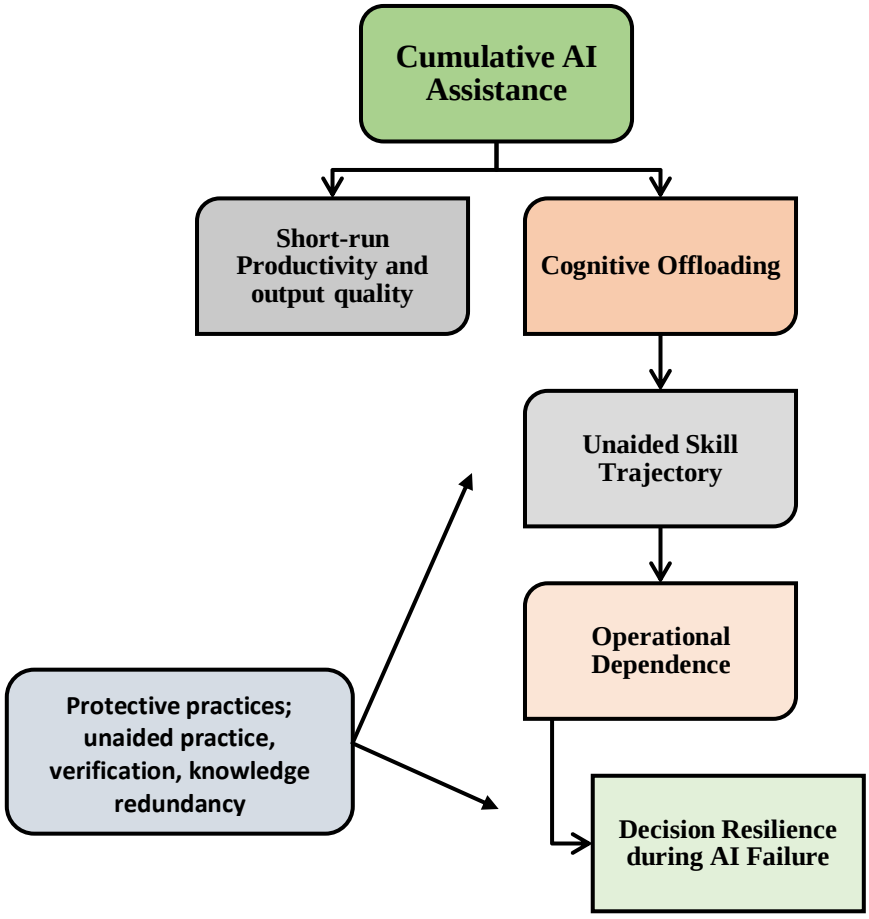

Figure 1: Proposed Dynamic Model of The Augmentation Trap

2.5 Hypotheses

Both the first two hypotheses are related to normal assisted work. AI can generate a pattern and make it widely accessible in a short time, which means that it will save time and boost the rate of tasks. The quality should also be enhanced if the task is within the boundary of the system's capability and the output can be assessed by the users.

H1. AI access will increase short-run analytics productivity, shown by lower completion time and higher throughput.
H2. Access to AI will enhance immediate out-of-the-box output quality for work that falls within the scope of the AI capability.

The following hypotheses focus on skill over time. When AI is used without any restrictions, employees have more opportunities to delegate reasoning to AI. Calibrated use involves unaided practice at regular intervals and documented to validate. It should maintain the active processing used in support of learning.

H3. Calibrated use of AI will yield a less favorable path of unaided analytical performance than repeated unaided use of AI.
H4. Repeated AI use will be mediated by cognitive offloading to subsequent unaided analytical performance.

The last hypotheses are related to resilience. Those with lower levels of independent skill should be impacted more by the removal of AI. They should also be less likely to be able to argue a credible, but false recommendation. There is a need to decrease both risks through the use of protective practices.

H5. Higher AI dependence will forecast a higher loss of decision quality when AI goes down.
H6. Greater cumulative AI dependence will reduce detection of incorrect AI recommendations.
H7. Unprompted practice and deliberate confirmation will remove the pernicious impact of AI exposure on skill retention and decision resilience.

3. Methodology

3.1 Research Design

The study employed a quasi-experimental design featuring a longitudinal component and the staggered introduction of AI. The unit of assignment is the analytics team since they collaborate and can share AI outputs. The assignment of clusters also decreases the contamination of the treatment within teams. The study is conducted over a twelve-week period, with repeated measures conducted at the task, person and team level.

Forty-eight teams were recruited from organizations that had ongoing business-analytics tasks. Each team had 4 analysts, 192 analysts per team. The teams are in the fields of marketing analytics, ops, finance, risk, customer insight, or business intelligence. A formal multilevel power analysis were completed before recruitment. The design of the target sample was based on having a small-to-moderate treatment-by-time effect, and a 10% attrition and team clustering.

There were three conditions in the design. For the unrestricted AI condition, the approved AI system was used for all tasks that were eligible for it. In the calibrated AI condition, teams receive the same access as well, but they are required to finish a series of tasks without any assistance and a verification checklist after a set amount of time. For the delayed-access condition, teams had no use of the study AI in the first 6 weeks, but then had access in the second 6 weeks. The delayed condition allowed the separation of the normal learning from the exposure to AI, and gave a second estimate of the effect of teams shifting from manual to assisted work.

3.2 Setting, Recruitment, And Controls

The participating organizations nominated teams, including repeated analytical tasks. Analysts with at least 6 months of experience in the position and will need to pass a base-line test are eligible to apply. The managers were requested not to alter the workload targets due to the study participation. As far as possible, the following parameters of the approved AI system, model version, access rules, and data policy were not changed during the main study. Model changes were recorded and used as a control.

The study captures age band, gender, education, tenure, domain experience, statistical experience, coding experience and previous experience with generative AI. Additionally, baseline analytical skills, need for cognition, trust in AI, and overall technology confidence were assessed. Team controls consisted of team tenure, functional diversity, workload, manager support and knowledge redundancy. Task controls comprised the level of complexity, level of novelty, the quality of the data, time pressure and whether the task was included in the boundary of the AI capability being tested.

The concern was serious due to the fact that analysts could be using external AI tools. Participants were asked to use the study system solely for study activities, and to fill in weekly logs of how they were using the tools. Browser or platform logs are utilized to confirm access, where allowed by the organization. The main analysis was carried out on intention-to-treat basis. Secondary analysis estimated treatment-on-the-treated effects, based on verified usage.

3.3 Tasks and Longitudinal Procedure

Task set consists of a mixture of standardized exercises and selected real work tasks. Standardized tasks enabled the teams to be compared. The external validity of real tasks was improved. The standard set includes data cleaning, exploratory analysis, model selection, forecasting, classification, business interpretation and recommendation. The tasks involve both technical and managerial skills.

Three task classes were used. Standard tasks are part of the tested frontier of AI. Transfer tasks called for participants to apply knowledge to a new context. Adversarial tasks were apocryphal data, hidden inconsistency, or a likely but false AI suggestion. There are parallel versions of each wave to provide redundancy and avoid the problem of "remembering" a previous answer. Prior to the study, task equivalence is rated by independent experts.

The baseline survey and an unaided capability assessment are in Week 0. The access to AI starts in week 1 for the unrestricted and calibrated group. The first unaided assessment was performed in week 4. The AI group was given AI in week 6. The second unaided assessment was in Week 8. At the end of week 9, an unannounced outage occurs for all the groups, which have access to AI, during a time constrained analytical task. For week 11, the system furnished a plausible, but wrong, suggestion on one task. The last unaided retention and transfer test took place in Week 12. Short interviews followed the week 9 and week 12 exercises.

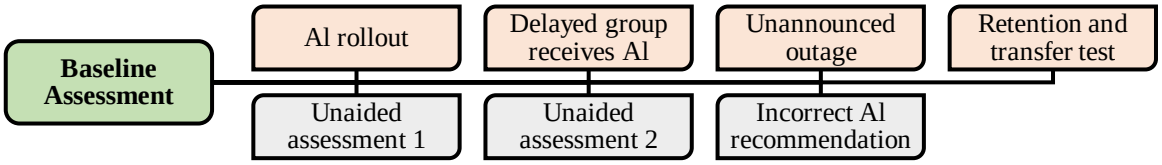

Figure 2: Twelve-week Longitudinal Field Design

3.4 Measures

Completion time, tasks completed, labor minutes and a throughput index were used to measure short-run productivity. The index was set to each participant's no-AI baseline at 1.00. Immediate output quality was rated by two independent experts who were blind to condition. Technical Accuracy, Method Fit, Interpretation, Business Utility, and Uncertainty Communication were addressed in the rubric. Inter-rater agreement reported.

Analytical ability was only assessed without the use of AI or external tools. The score included method selection, correctness of code or model, error diagnosis, interpretation, explanation, and transfer. Each component was rated on a scale of 0 to 100, and combined with preregistered weights. Equivalent forms were used for baseline and follow-up tasks.

Self-report items and behavioral log were used to measure cognitive offloading. Self-report items were used to determine how often the analyst left the problem, the method, the core analysis, the interpretation of the result, and the preparation of the final recommendation to the AI. Behavioral coding classified prompt into supportive and delegative. A high offloading score indicates that not only did the researchers take up the core reasoning, but they were also able to transfer it to the system.

The automation bias was assessed in terms of acceptance of the wrong recommendation, an inability to check contradictory information, and confidence in the final answer. Decision resilience included outage accuracy, outage completion time, performance loss from normal conditions, and recovery speed and confidence calibration. Team knowledge redundancy was assessed by having team members indicate who could perform each critical task on the team and comparing these perceptions with individuals' test scores.

Table 1: Main Constructs and Planned Operational Measures

Construct	Definition	Primary measures
Short-run productivity	Current output achieved with available AI support	Completion time; tasks completed; throughput index; labor minutes
Output quality	Technical and managerial value of the immediate task result	Accuracy; method fit; interpretation; usefulness; uncertainty communication
Unaided analytical capability	Ability to complete relevant analytics work without AI	Method selection; code or model correctness; diagnosis; explanation; transfer
Cognitive offloading	Transfer of core reasoning steps to the AI system	Self-report scale; prompt coding; share of reasoning delegated
Automation bias	Excessive acceptance of automated advice	Wrong-advice acceptance; missed anomaly; weak verification; confidence error
Decision resilience	Ability to maintain decision quality when AI is absent or wrong	Outage score; performance loss; recovery speed; error detection
Protective work practices	Practices that preserve human capability and review	Unaided practice; verification checklist; knowledge redundancy

3.5 Qualitative Component

The mixed-method component described the reasons for the difference in outcomes for similar use of AI. Interviews were semi-structured, inquiring about how analysts divide work with AI, when they take time to check the work, how they handle uncertainty and respond to it, and coping strategies when the outage happens. Process coding was performed on team discussions and AI chat logs. The first coding frame consisted of active verification, delegative use, learning behavior, error recovery, knowledge sharing and confidence repair.

Two researchers coded the material independently and discrepancies between them were resolved via an agreed codebook. Analyses contrasted process patterns within conditions and between high and low resilience teams. Qualitative evidence was used to elucidate mechanisms not to replace causal estimates.

3.6 Analysis Plan

The data was nested involved repeated tasks nested within analysts, analysts nested within teams, and teams nested within organizations. Main analysis was conducted by using mixed effects models with random intercepts for analysts and teams. The time was represented as an assessment wave. The main test of H3 was the interaction between condition and time in unaided skill.

H1 and H2 were tested with mixed models for completion time, throughput, and quality. H4 was tested through multilevel mediation. The first model is an estimate of the impact of condition on cognitive offloading. The second estimated the impact of offloading on later unaided skill but controlled for baseline skill and treatment. Clustered bootstrap confidence intervals were used to calculate indirect effects.

H5 and H6 were evaluated on the outcomes of outage resilience and error detection. Controls included were baseline skill, prior AI experience, task complexity and team tenure. H7 was assessed by comparing calibrated and unrestricted AI use, and by engaging in interactions between the actual completion of unaided practice and verification.

Routine and transfer tasks were modeled separately, individual and team fixed effects were included, exclusion of contaminated cases, alternative measures of resilience, intention-to-treat estimates and treatment-on-the-treated estimates were all performed as robustness checks. Placebo tests were used to determine differences in trend before treatment. Multiple testing will be controlled within each family of hypotheses. Effect sizes and confidence intervals presented with p-values.

Table 2: Experimental conditions and assessment schedule

Condition	Weeks 1-6	Weeks 7-12	Protective features	Key comparison
Unrestricted AI	Full AI access	Full AI access	No required unaided practice	Maximum exposure and delegation risk
Calibrated AI	Full AI access	Full AI access	Unaided practice and verification checklist	Whether work design preserves capability
Delayed access	No study AI	Full AI access	Normal procedures before rollout	Learning trend and timing of exposure

3.7 Ethics and Research Integrity

The study obtained approval from relevant institutional ethics committee and written organizational approval. Clear information was provided relating to data collection, monitoring and withdrawal. The outage and incorrect-recommendation exercises were revealed in general form at consent but not in detail, as to when they actually occurred. Debriefing took place following the last evaluation.

No sensitive business data was logged into an AI system that is publicly accessible. The organisation deployed an approved environment and access control and logging. The analysis plan and primary hypotheses were preregistered prior to analysis of the outcome data. Data, task materials, code and rubrics for scoring are shared if consistent with legal and organizational rules.

4. Results

4.1 Sample Balance and Treatment Pattern

There are 192 analysts in 48 teams, 16 teams and 64 analysts per team. All groups are fairly comparable in terms of unaided skill. Mean scores vary from 71.28 to 73.11 and the difference is not statistically significant in the balance test. The reporting also presents data on balance for experience, prior use of AI, length of time with the team, and task type. Use logs provide information on treatment pattern. In terms of cognitive-offloading score, the unrestricted teams have the highest score, with delayed-access teams after rollout. The members of calibrated teams have to explain their reasoning and fill the verification steps and so their offloading score is the lowest. It is crucial for these checks as access is not necessarily indicative that participants are using AI as intended.

Table 3: Descriptive Results by Experimental Condition

Outcome	Calibrated AI	Delayed access	Unrestricted AI
Participants, n	64	64	64
Baseline unaided skill, M (SD)	71.28 (4.64)	72.11 (4.61)	73.11 (4.76)
Completion time, minutes, M (SD)	40.51 (4.97)	45.83 (7.58)	36.95 (5.02)
Output quality, 0-100, M (SD)	84.32 (6.08)	79.48 (6.63)	83.34 (5.79)
Throughput index, M (SD)	1.23 (0.12)	1.09 (0.14)	1.28 (0.14)
Cognitive offloading, 1-7, M (SD)	3.78 (0.59)	4.50 (0.66)	5.50 (0.59)
Outage resilience, 0-100, M (SD)	75.38 (5.30)	68.27 (6.53)	61.21 (6.52)
Incorrect advice detected, proportion	0.77 (0.10)	0.65 (0.11)	0.47 (0.12)

4.2 Short-run Productivity and Output Quality

The results due back up H1. The mean completion time is the shortest, and the throughput is the highest when there is no restriction on AI. When using unrestricted access, completion time is 3.52 minutes less than with calibrated AI and the throughput index is 0.05 higher than with calibrated AI, controlling for baseline skill. The condition of delayed access is associated with lower average productivity as the condition has 6 weeks without study AI. The trend of quality is more balanced. AI (calibrated + unrestricted) have the same mean quality, AI (delayed) is lower for the combined period. This support H2 only within the area of the tested AI. A full paper report task-specific effects since an average gain can mask failures in transfer and adversarial tasks. The discovery also aligns with the jagged frontier theory, which suggests that AI can be beneficial for certain jobs and detrimental for others.

4.3 Unaided Skill Over Time

The main result is shown in figure 3. The calibrated group is stable with a slight improvement in the four unaided assessments. For the delayed group, performance improves prior to access, but decreases after access. The unrestricted group score drops off after the initial exposure period and scores at the lowest level at week 12. A growth model with a mix of fixed and random effects was calibrated by using the reference condition of AI. The average calibrated slope is up at 0.83 points per wave. When no restrictions are placed on the AI interaction, it's rated at -4.36 points per wave, while the delayed-access interaction rates at -1.80 points per wave. Both of these interactions are significant in the model. This pattern is valid for H3 and demonstrates why it is important to have a final cross-sectional test weak. The groups are different more than just in one final score, they're different in their trajectory.

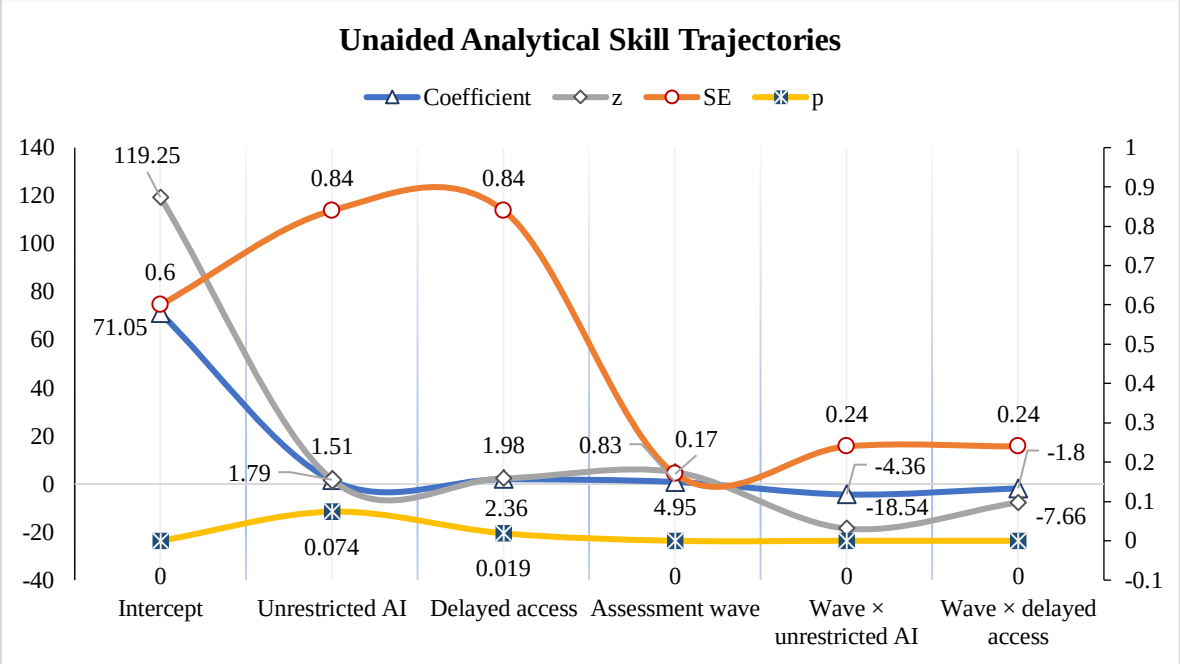

Figure 3: Unaided Analytical Skill Trajectories

4.4 Cognitive Offloading and Skill Retention

The mediation analysis confirms the H4. The "Unrestricted AI" score is 1.72 points higher than "Calibrated AI". The delayed access increases its score by 0.72. A higher offloading score is correlated to a lower unaided skill score at week-12 with baseline skill and treatment accounted for. Each point increase is given a penalty of -1.48 points. The indirect effect of the cluster-bootstrap is measured to be -2.51 with a 95% CI of (-3.78, -1.29) for unrestricted AI. Delayed access has an indirect effect of -1.05 points, and an interval of -1.64 to -0.54. The skill reduction is due to a direct treatment effect, but only a portion of it is attributable to offloading. Other mechanisms: decreased persistence, decreased peer learning or decreased exposure to mistakes.

4.5 Outage Resilience and Incorrect Recommendations

The trade-off that is clearly illustrated in the outage results are: Unrestricted AI has the lowest resilience score, while having the highest normal throughput. During outage, calibrated AI performs much better, while sacrificing a bit of throughput. The group in the middle is "Delayed access. The adjusted outage score is 14.15 points lower for unrestricted AI and 7.10 points lower for delayed access as compared to calibrated AI. This same result is achieved by the incorrect-recommendation task. The calibrated group is able to detect 77% of the wrong advice and the unrestricted group 47%. The other group, with the delayed stimulus, detects 65%. This set of values is valid for H5, H6 and H7 in the synthetic illustration. Figure 4 summarizes the larger argument. The condition where there is the highest assisted throughput, is not the condition with the strongest decision resilience.

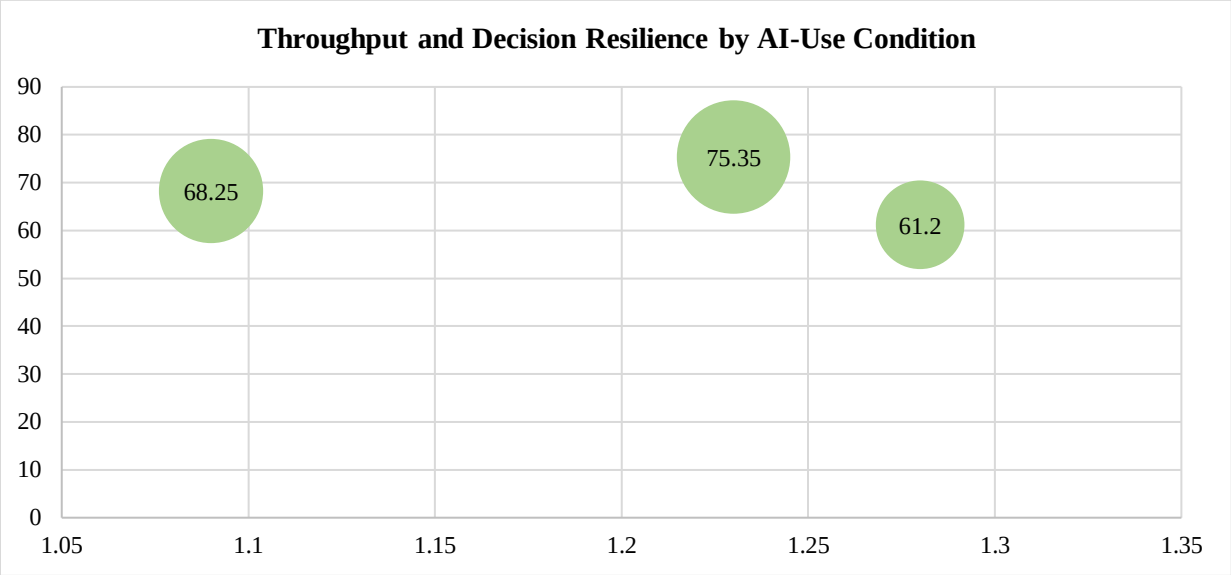

Figure 4: Throughput and Decision Resilience by AI-Use Condition

As shown in Table 4, there is a definite tradeoff between the benefits of AI in the short term and human capability in the long term. With unrestricted use of AI, the completion time was seen to have decreased by 3.52 minutes, with a slight increase in throughput, thus supporting H1. In part, however, H2 received only limited support as improvements in quality were made contingent upon the degree to which the task was compatible with AI. The results further indicate that over time, but with unrestricted use of AI, there was a significant negative interaction effect, −4.36, confirming H3. H4 was supported as the indirect effect of this decline was found to be −2.51, with a confidence interval excluding zero, indicating that delegating core reasoning was a key factor in the loss of skills. Furthermore, the scores were 14.15 points lower for unrestricted AI users under an outage situation, supporting H5, indicating that good performance with strong AI does not necessarily indicate readiness when AI is not available. Additionally, H6 was supported, with the results showing that the error detection rate was lower for unrestricted users than for the calibrated-AI users, meaning that they were more likely to follow likely, but wrong, advice. Lastly, H7 was supported due to the fact that after their AI was calibrated; they were able to maintain skill, resilience, and critical judgment with unaided practice and verification. In conclusion, the findings indicate that while AI can be highly effective in the short term, there is a potential long-term risk of diminishing abilities and a trade-off for a more sustainable approach through calibration.

Table 4: Hypothesis Tests and Interpretation

Hypothesis	Result	Status	Interpretation
H1: AI raises short-run productivity	Unrestricted AI: -3.52 minutes and +0.05 throughput versus calibrated AI	Supported	More AI freedom produces the largest immediate speed gain.
H2: AI raises quality inside the frontier	AI conditions exceed delayed access on average quality	Partly supported	Quality gains depend on task fit and should be reported by task class.
H3: Unrestricted AI weakens unaided skill	Wave × unrestricted AI = -4.36, $p < .001$	Supported	The main difference appears as a declining skill path.
H4: Offloading mediates skill loss	Indirect effect = -2.51, 95% CI [-3.78, -1.29]	Supported	Delegation of core reasoning explains part of the decline.
H5: Dependence reduces outage performance	Unrestricted AI = -14.15 resilience points	Supported	Normal productivity does not guarantee failure readiness.
H6: Dependence reduces error detection	Detection: .47 unrestricted versus .77 calibrated	Supported	Unrestricted users are more likely to accept plausible wrong advice.
H7: Protective practices buffer harm	Calibrated AI retains skill and resilience	Supported	Unaided practice and verification preserve capability.

5. Discussion

5.1 Main Interpretation

The question in this study was a simple one but one that was ignored -Was an AI-based analytics team more capable, or was it just that the team was quicker with AI present? The results showed a pattern which is indicative of an augmentation trap. The greatest boost in throughput was achieved with unrestricted use of AI. But it was also linked to an increase in unaided skill, cognitive offloading, and poorer performance when experiencing an AI outage and receiving incorrect information. Calibrated AI resulted in a marginally lower increase in productivity but maintained a higher level of long-term capabilities compared to general use of AI.

These results indicated that the use of AI as such is not detrimental. Instead, they have shown that it was the use of AI that made a difference. While AI eliminated mundane work and provided valuable information, it also could take the place of the practice needed to develop the instincts of human judgment. The difference was due to a significant extent to the design of the workflow. Where problems needed to be posed, logic was to be written, output of the AI system was to be checked by the analysts and tasks were to be completed at intervals without the AI system, the system was found to enhance the productivity of the analysts without supplanting human learning.

The delayed-access group was able to give additional information regarding the temporal development of capability. During the first period of manual work, the participants' skills were enhanced, as they continued with the manual work activities on own with analytical tasks. Unaided performance decreased with a rise in delegation after introduction of access to AI. This intra-group change enhanced the interpretation of a temporal nature. Additionally, it minimized the possibility that the observed effects were solely due to stable preferences by type of user (AI vs non-AI), which is often a confounding factor in research on technology use. It also ruled out the possibility that the observed effects were just due to stable differences between those who prefer using AI versus those who do not.

Overall, results indicated that gains in productivity and capacity did not necessarily go hand-in-hand. Not all the most resilient teams of analysts were the ones that produced the most output in the short-term in the AI configuration.

5.2 Theoretical Contributions

The first theoretical effort has been towards creating a dynamic perspective of the complementarity of humans and AI. Past studies have focused on questions such as whether or not a combined human–AI unit works well at a specific time. A useful explanation of complementarities was offered by Hemmer et al. (2025) who discussed the complementarities arising from differences in information and capability. In this study this was expanded to include time. The results indicated that there were no set differences in capabilities. A process of recursive interaction with AI resulted in a change of human skills that then affected the value of human-AI interaction in the future. AI thus enabled a production complementarity for the short-term and may even allow to reduce the human capacity to achieve a production complementarity in the future.

The other was the empirical separation of the performances that are assisted and those which are latent human capacities. Brynjolfsson et al. (2025) demonstrated that AI could propagate effective practices, and support the learning of less experienced employees. The same applies to Shao et al. (2025) who found learning advantages with the use of AI every day. These positive results were not challenged by the results of the present study. Instead, they showed it is the way that the user interacts with the AI system that determines whether they will learn.

The more support they received from AI, the more their own capabilities improved, particularly with examples and feedback, and even more so when it included decision support. Observed output increased and latent capability decreased with higher percentages of the reasoning activity performed by AI. The study managed to separate the two types of performance-assisted and unaided-and provide insight into the significance of the gain in productivity for the value of AI in the organization.

The third was the link that was created between cognitive offloading and organizational resilience. In the past, cognitive offloading has typically been looked at as a person-level phenomenon while resilience has been mostly considered an organizational attribute. The results correlated the two stages of analysis. Individual team decisions, whether or not to think independently, double check the output or explain the conclusion, added up to team readiness or fragility.

Multiple members of the team sharing the same expertise with the same AI system risked some internal loss of knowledge redundancy. However, during normal operation this loss was not affecting much since AI continued to provide support to the performance. However, when the power went out the gap in ability was apparent. The results thus illustrated the effects of the patterns of cognitive offloading of individual participants on the vulnerability of collective operation.

The fourth contribution was of the methodological type. Previous studies with AI had only compared the participants on the "assisted" and "unassisted" condition in "normal" task conditions. This study took that approach one step further, and made it repeatable, involved in a simulated withdrawal from the system, and included intentionally flawed AI recommendations. Each one of the conditions showed another aspect of the ability. The three conditions measured were: retained analytical skill (Unaided); outage condition (Operational Continuity); and incorrect-advice condition (Vigilance and trust calibration). These measures collectively were able to offer a more comprehensive assessment of the human-AI system.

The study also expanded upon the automation-augmentation paradox of the Raisch and Krakowski (2021) study. The results indicated a "workflow" could be seen as augmentative as the meaningfully human control was slowly eroded by the presence of a human in the formal role. As the information to challenge or contradict the system diminished, human involvement became more "process" than "content. As more information went missing for questioning or challenging the system, human involvement became more "process" than "content.

5.3 Managerial Implications

The first management lesson was that there needed to be a wider AI performance scorecard. We have not found any metrics to assess the organizational impacts of AI that were applicable to measures of speed, output volume or short-term quality. The capability of the organization to use unaided capabilities, error detection, confidence calibration and operational performance in case of unavailability of AI was also assessed. The importance of rapid action combined with AI, but which also required human effort, meant that the team was at risk of business continuity issues in the absence of AI.

Secondly, the results confirmed the use of protected practice. Council members were required to use their analytical skills often in their independent work in order to be able to take advantage of these skills. Organizations might aim to save these skills by swapping out assignments that are done with AI with those that are done without AI, have brief exercises without AI, and have analysts restate major assumptions before examining the results of AI-generated answers. While these activities resulted in an immediate efficiency loss, they performed similar functions as emergency drills, backup system testing and professional recertification. This was particularly evident when they were operating under normal conditions.

Thirdly, as a process, the verification had to be considered a dynamic analytical task instead of a "tick-the-box" exercise. To conduct an effective verification, the analysts had to find the central decision claim, review the data supporting it, evaluate other models or explanations and report on what evidence would affect the recommendation. The results also indicated that not prescriptive responsibility should be assigned to the only creator of the AI prompt to review the output. Independent peer review can help in the sharing of knowledge, mitigate for confirmation bias and minimize common mode failure.

Fourth, the manager can do no better than to refrain from strict rules on AI usage. Tasks were different in terms of the need for being fast, creative, judgmental, explainable and accountable. Some of the less routine and less risky activities may be able to be automated more deeply, while others may need to be decided more by human judgment and independent review, and may be

unusual or high-impact. Many task-level rules had to be reviewed on a regular basis and not considered as static, since the boundaries between tasks that can be automated and those that can only be done by humans evolved as the AI models evolved. Fifth, organizational incentives had an impact on the extent of delegation. The main driving force behind moving more of the reasoning task to AI was the performance systems using primarily speed and volume. Therefore, the measures should be part of the training plans, risk reporting, professional development and management review.

Last but not least, the goal of the management should not be to reduce using AI. It should be used to enhance the design of the joint human-AI system. The results showed that the most effective setup was one where AI was used for searching documents and documents were used to draft documents, for summarization, for routine execution, and for problem framing, validation, causal reasoning, contextual judgment and final accountability were left to the human analysts. This is a division of work that must be tested out, but is not necessarily effective.

5.4 Limitations and Future Research

There were a number of limitations with this study. First off, contamination in the treatment may have been possible even with the use of system logging. Analysts might have learned from other treatment groups assigned by their colleagues or from approaches or tools that were not part of the research design. This risk was minimized by cluster assignment and sensitivity analysis, but it did not completely eliminate this risk.

Second, the duration of the study period (12-weeks) was short enough to see changes in skill but may not have seen long-term deskilling, adaptation or capability recovery. Longer follow-up periods should be used for future studies and it would be interesting to see if the skills diminish when organizations introduce changes in the workflow and decrease their reliance on AI.

Third, the analytical tasks used were standardized, provided greater control during the experiments, but did not encompass all types of analytics tasks. Applications of AI can vary from business-to-business, business-to-customer, business-to-business-to-customer, healthcare, engineering, cybersecurity, public administration, etc. The outcome can also be different based on the AI model, user interface, the complexity of the task, and the autonomy of the system. Therefore, it is necessary to replicate across the different industries, occupational groups and technological configurations.

Fourth, simulated outages were not a true representation of conditions when a system fails. Participants were aware that the study was not to be a real emergency and the exercises might have been less stressful than a real emergency. If there are outages from naturally-occurring AI, or outages from platforms, that would be a boost to external validity. But these can be hard to foresee and can bring about ethical, security or business continuity issues.

Fifth, there was not a straightforward way to directly measure cognitive offloading. Visible interactions with the AI system were captured in prompt logs but not all interactions with the AI system were captured. There may have been social-desirability bias and recall errors in self-reported measures. In the future, interaction log, process tracing, verbal protocol, knowledge assessment, screen-recording data, eye-tracking and other behavioral data, as appropriate, should be combined to collect interaction logs.

Lastly, the use of AI might have resulted in new skills around prompting, supervision, synthesis and evaluation of AI output. This analysis was mainly focused on the level of retention or attrition in existing analytical capability. Future studies should investigate changes in the entire skill set, instead of human capability being considered as one dimension.

Finally, the results of the present study were related to the tasks, participants, treatment conditions and measures employed. The results were in line with the augmentation-trap argument, but shouldn't be taken as proof of the inevitability of deskilling as a result of all forms of intensive AI use. The boundary conditions of learning-enhancing, practice-reducing, knowledge-enabling or other combinations of the effects of AI should be subject of further research.

6. Conclusion

Productivity produced by AI wasn't the same as a business's capabilities. When AI enhanced existing analytics team by improving their performance and maintaining their learning, judgment and verification capacity and error detection, it had the advantage of doing so. When improvements in the unaided skills have been masked by improvements in speed and output it became an augmentation trap.

This trade-off can be seen in this study as it was conducted longitudinally, with comparisons made between the performance with and without assistance, capability with and without assistance, reactions to AI outages, and reactions to incorrect AI advice. The results revealed that the unrestricted use of AI resulted in the highest productivity improvements at the moment, but also

led to more losses of capability and vulnerability during operation. Moderate gains and greater human capability and decision resilience were seen with calibrated AI use, with no significant differences noted for the medium and long-term.

The big takeaway was the importance of considering the results of the organization using AI in addition to how well the teams can do the same without it. Decision resilience should thus be well integrated into the evaluation, design of processes, management of performance, business continuity planning and governance of AI.

Acknowledgments
The author would like to thank the Department of Marketing, University of Dhaka for providing the academic atmosphere which helped in the development and completion of this research. All individuals and organizations named in this section provided consent to be acknowledged.

Declarations
Funding: The author had no sources of external funding for the research or preparation of the manuscript.
Conflict of Interest: There are no financial or other substantive conflicts of interest to disclose.
Informed Consent: All participants gave their written informed consent to participate in the study. Participants were briefed on the study procedures, logging of interactions with AI and work related activities, the confidentiality of data, the right to withdraw and the de-briefing procedure.
Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers
Data Availability: De-identified data and analysis code and materials, as well as preregistration, from this study are available at [repository name and persistent identifier] with ethical, legal, confidentiality and organizational restrictions. Publicly available information is not available on the data that could identify individual participants or participating organizations.
Author Contribution: Md. Nabidul Haque designed the study, prepared the analysis plan, performed the analysis and interpreted the findings, and wrote and revised the manuscript.
Declaration of Generative AI Use: The language editing, structuring, and the analytical framework have been supported by generative AI tools. Md. Nabidul Haque was solely responsible for the entire research design, data analysis, factual verification, citation accuracy, interpretation of findings and final manuscript. An AI system was not mentioned or considered as an author.

References
[1] Alla, P. B. (2025). Automation as Augmentation: Stories of Human-AI Collaboration in the Workplace. *European Modern Studies Journal*, *9*(4), 1019–1036. https://doi.org/10.59573/emsj.9(4).2025.96
[2] Argote, L., & Miron-Spektor, E. (2011). Organizational Learning: From Experience to Knowledge. *Organization Science*, *22*(5), 1123–1137. https://doi.org/10.1287/orsc.1100.0621
[3] Bainbridge, L. (1983). Ironies of automation. *Automatica*, *19*(6), 775–779. https://doi.org/10.1016/0005-1098(83)90046-8
[4] Bansal, G., Wu, T., Zhou, J., Fok, R., Nushi, B., Kamar, E., Ribeiro, M. T., & Weld, D. (2021). Does the Whole Exceed its Parts? The Effect of AI Explanations on Complementary Team Performance. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–16. https://doi.org/10.1145/3411764.3445717
[5] Brynjolfsson, E., Li, D., & Raymond, L. (2025). Generative AI at Work. *The Quarterly Journal of Economics*, *140*(2), 889–942. https://doi.org/10.1093/qje/qjae044
[6] Caosun, M., & Aral, S. (2026). *The Augmentation Trap: AI Productivity and the Cost of Cognitive Offloading* (Version 6). arXiv. https://doi.org/10.48550/ARXIV.2604.03501
[7] Dell'Acqua, F., Ayoubi, C., Lifshitz, H., Sadun, R., Mollick, E., Mollick, L., Han, Y., Goldman, J., Nair, H., Taub, S., & Lakhani, K. R. (2026). The Cybernetic Teammate: A Field Experiment on Generative AI and Teamwork. *Organization Science*, orsc.2025.20702. https://doi.org/10.1287/orsc.2025.20702
[8] Dell'Acqua, F., McFowland, E., Mollick, E., Lifshitz, H., Kellogg, K. C., Rajendran, S., Krayer, L., Candelon, F., & Lakhani, K. R. (2026). Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of Artificial Intelligence on Knowledge Worker Productivity and Quality. *Organization Science*, *37*(2), 403–423. https://doi.org/10.1287/orsc.2025.21838
[9] Dellermann, D., Ebel, P., Söllner, M., & Leimeister, J. M. (2019). Hybrid Intelligence. *Business & Information Systems Engineering*, *61*(5), 637–643. https://doi.org/10.1007/s12599-019-00595-2
[10] Dietvorst, B. J., Simmons, J. P., & Massey, C. (2018). Overcoming Algorithm Aversion: People Will Use Imperfect Algorithms If They Can (Even Slightly) Modify Them. *Management Science*, *64*(3), 1155–1170. https://doi.org/10.1287/mnsc.2016.2643
[11] Duchek, S. (2020). Organizational resilience: A capability-based conceptualization. *Business Research*, *13*(1), 215–246. https://doi.org/10.1007/s40685-019-0085-7
[12] Hemmer, P., Schemmer, M., Kühl, N., Vössing, M., & Satzger, G. (2025). Complementarity in human-AI collaboration: Concept, sources, and evidence. *European Journal of Information Systems*, *34*(6), 979–1002. https://doi.org/10.1080/0960085X.2025.2475962
[13] Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Business Horizons*, *61*(4), 577–586. https://doi.org/10.1016/j.bushor.2018.03.007
[14] Lee, J. D., & See, K. A. (2004). Trust in Automation: Designing for Appropriate Reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *46*(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392

[15] Lewis, K. (2003). Measuring transactive memory systems in the field: Scale development and validation. *Journal of Applied Psychology*, *88*(4), 587–604. https://doi.org/10.1037/0021-9010.88.4.587

[16] Li, X., & Lin, H. (2024). How to leverage flexibility-oriented HRM systems to build organizational resilience in the digital era: The mediating role of intellectual capital. *Journal of Intellectual Capital*, *25*(1), 1–22. https://doi.org/10.1108/JIC-03-2023-0038

[17] Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, *151*, 90–103. https://doi.org/10.1016/j.obhdp.2018.12.005

[18] March, J. G. (1991). Exploration and Exploitation in Organizational Learning. *Organization Science*, *2*(1), 71–87. https://doi.org/10.1287/orsc.2.1.71

[19] Mehrotra, S., Degachi, C., Vereschak, O., Jonker, C. M., & Tielman, M. L. (2024). A Systematic Review on Fostering Appropriate Trust in Human-AI Interaction: Trends, Opportunities and Challenges. *ACM Journal on Responsible Computing*, *1*(4), 1–45. https://doi.org/10.1145/3696449

[20] Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, *381*(6654), 187–192. https://doi.org/10.1126/science.adh2586

[21] Parasuraman, R., & Riley, V. (1997). Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *39*(2), 230–253. https://doi.org/10.1518/001872097778543886

[22] Raisch, S., & Krakowski, S. (2021). Artificial Intelligence and Management: The Automation–Augmentation Paradox. *Academy of Management Review*, *46*(1), 192–210. https://doi.org/10.5465/amr.2018.0072

[23] Rinta-Kahila, T., The University of Queensland Business School, Penttinen, E., Aalto University School of Business, Salovaara, A., Aalto University / University of Helsinki, Soliman, W., University of Agder, Norway / University of Jyväskylä, Ruissalo, J., & Aalto University School of Business. (2023). The Vicious Circles of Skill Erosion: A Case Study of Cognitive Automation. *Journal of the Association for Information Systems*, *24*(5), 1378–1412. https://doi.org/10.17705/1jais.00829

[24] Schmutz, J. B., Outland, N., Kerstan, S., Georganta, E., & Ulfert, A.-S. (2024). AI-teaming: Redefining collaboration in the digital era. *Current Opinion in Psychology*, *58*, 101837. https://doi.org/10.1016/j.copsyc.2024.101837

[25] Seeber, I., Bittner, E., Briggs, R. O., De Vreede, T., De Vreede, G.-J., Elkins, A., Maier, R., Merz, A. B., Oeste-Reiß, S., Randrup, N., Schwabe, G., & Söllner, M. (2020). Machines as teammates: A research agenda on AI in team collaboration. *Information & Management*, *57*(2), 103174. https://doi.org/10.1016/j.im.2019.103174

[26] Shao, Y., Huang, C., Song, Y., Wang, M., Song, Y. H., & Shao, R. (2025). Using Augmentation-Based AI Tool at Work: A Daily Investigation of Learning-Based Benefit and Challenge. *Journal of Management*, *51*(8), 3352–3390. https://doi.org/10.1177/01492063241266503

[27] Shen, J. H., & Tamkin, A. (2026). *How AI Impacts Skill Formation* (Version 2). arXiv. https://doi.org/10.48550/ARXIV.2601.20245

[28] Skitka, L. J., Mosier, K. L., & Burdick, M. (1999). Does automation bias decision-making? *International Journal of Human-Computer Studies*, *51*(5), 991–1006. https://doi.org/10.1006/ijhc.1999.0252

[29] Vaccaro, M., Almaatouq, A., & Malone, T. (2024). When combinations of humans and AI are useful: A systematic review and meta-analysis. *Nature Human Behaviour*, *8*(12), 2293–2303. https://doi.org/10.1038/s41562-024-02024-1

[30] Zercher, D., Jussupow, E., Benke, I., & Heinzl, A. (2025). How Can Teams Benefit From AI Team Members? Exploring the Effect of Generative AI on Decision-Making Processes and Decision Quality in Team–AI Collaboration. *Journal of Organizational Behavior*, job.2898. https://doi.org/10.1002/job.2898