From Automation to Augmentation: How Expertise Complementarity Converts Generative AI into Team-Level Productivity

Abdullah Mohammad Yeaqubi
Correspondence: abdyeaqubi@gmail.com

In progress 2026

ABSTRACT

Generative artificial intelligence has diffused rapidly across knowledge work, yet adoption does not reliably translate into collective performance. This study investigates how generative AI usage and team productivity are connected through the lens of expertise complementarity. This cross-sectional secondary analysis combines the results of published findings from the 2025 Stack Overflow Developer Survey, based on the responses of 11,202 individuals from the workflow module, distributed across the globe, with comparisons based on constructs of adoption, sentiment, trust, task boundaries, productivity, and collaboration. Results show an augmentation gap, with 84% using or planning to use AI tools and positive sentiment dropping to 60%, while positive sentiment in terms of team collaboration among AI user agents drops to 17% (52% report positive productivity effects). Concerning accuracy, it reaches 87%, and 76% do not want to deploy or monitor AI. Value is added by expertize by selectively delegating, verifying, and reintegrating, not by automating everything. The study presents an expertise-delegation-integration framework and outlines governance practices that are able to maintain quality while ensuring that individual speed is transferred to long-lasting team performance within organizations and settings.

Keywords: Generative AI; Expertise Complementarity; Team Productivity; Task Delegation; Digital Transformation

1. Introduction

Generative artificial intelligence (GenAI) is going beyond the realm of mere experimentation and becoming a regular part of knowledge work. Such research, conducted with control groups, has been correlated with improved completion time and higher assessment scores on writing intensive tasks (Noy & Zhang, 2023), and in the field with customer support associates, productivity gains resulted from using AI focused primarily on less experienced workers (Brynjolfsson et al., 2025). The same can be said for creative problem-solving experiments, which also are reported to be more novel or useful under certain human–AI configurations (Boussioux et al., 2024; Doshi & Hauser, 2024). The results provide a rationale for the excitement over enterprise adoption, but do not constitute proof that enterprise team productivity. Teams need to synchronize interdependent activities, combine disparate results, maintain synchrony of understanding, and identify errors that are not so conspicuous that they are obvious to anyone.

The actual question is not if can GenAI generate text, code, analysis, or ideas faster. It is how organizations distribute the responsibility of work among humans and AI without compromising on judgment, accountability, and coordination. While algorithmic output replaces human effort in the process of automation, algorithmic output and human effort are combined in the process of augmentation to play different roles. According to the automation-augmentation paradox, there will be both substitution and complementation not one or the other as automation of labor progresses (Raisch & Krakowski, 2021). The role-based study of machine teammates also addresses roles and interaction protocols as well as mutual adjustment (Seeber et al., 2020). However, in many implementations, access to the tool is the intervention and frequency of access is the outcome, and the process of converting the adoption of the tool to the collective performance isn't theoretically specified.

This separation has an economic impact. If the task can be done faster locally, but the correction, approving, and coordinating do not escalate to the colleagues, a measured effort goes down. If the migration is not documented, then the dashboards that are created overestimate the productivity and underestimate the risk concentration. To analyse the team level, however, a wider production scope is needed, which implies the consideration of all the stages from prompt production to accepted, deployable, and shared output.

That process has at its core expertise. Advanced practitioners are able to anticipate domain constraints, develop probes, identify unlikely output, and link up generated material with local routines. Novices may benefit from AI-generated missing templates and procedural knowledge (Brynjolfsson et al., 2025), but lack of expertise diminishes the ability to detect errors and to adapt to context. It is essential for human–AI symbiosis to rely on task conditions and distributed judgment (Jarrahi, 2018). Algorithmic systems also redefine the act of knowing in practice as validation, exception management and responsibility moves from one occupation to another. The value of an organization can thus

be understood as the result of a sociotechnical configuration and not only of the capability of a model (Faraj et al., 2018).

Complementarity research shows that this is only true when the pattern of errors is different and the allocation rule suggests allocating the task to a different contributor. This result is jointly determined by users' skills and algorithmic adjustments (Hemmer et al., 2025). Another key aspect of GenAI is the introduction of interaction costs, such as iterations of the prompt, source checking, debugging, and repair, which are often not tracked by traditional productivity measures (Simkute et al., 2024). In addition, knowledge-management analysis also suggests that too much delegation can lead to a decrease in experiential learning and rejuvenation of tacit knowledge. The team that gets the work done quickly will experience lost quality, multiple checks, ownership fragmentation or capability erosion, but still be able to see the benefits at the individual level (Alavi et al., 2024).

The previous investigations have three drawbacks. First, experimental studies typically research the individual task of a limited range of jobs, instead of team routines that rely on the interdependence of these tasks. Second, research in fields is likely to obtain average results of worker effects, and seldom brings together the expertise, task allocation, trust, and collaboration into one model. Third, organisational structures make a conceptual separation between automation and augmentation, but do not provide concrete indicators of when they have crossed the line (Berente et al., 2021; Hillebrand et al., 2025; Raisch & Krakowski, 2021). Also considered is the limited duration of the studies for available GenAI, where quality and team integration are considered comparatively weak. A cross-sectional diagnostic will not be able to determine long run causality but will help to test if the adoption patterns of the present are consistent with the suggested mechanism.

The study seeks to shed light on how GenAI use is translated into team level productivity without compromising on quality. The study's goal is to provide an understanding of how the use of GenAI can lead to team-level productivity without sacrificing quality. The main goals are to measure the gaps between adoption and positive experiences, to determine if there are similar productivity gains when working collaboratively, to find patterns of expertise-driven trust and verification, and to find out where the task boundaries and risks lie that affect augmentation. These goals view team productivity as a multi-phase process that includes efficient execution, coordinated integration, proper delegation, and quality retention instead of uptake of tools.

This framing separates organizational efforts (validation, combination, communication, and acceptance) that are less apparent but essential to ensure generated contributions are accepted across interdependent roles from visible acceleration. It also makes quality preservation a part of productivity, but not the part that comes after it. This contribution is triple. In theory, this study combines the concept of expertise complementarity, task–technology fit, and sociotechnical coordination and presents a single conversion model that explains how the same access results in different organizational outcomes. Methodologically, it illustrates the use of public cross-sectional indicators to support an open and clear (though noncausal) diagnostic process, while allowing for the inclusion of item-specific denominators, separation of adjacent measures, and making no individual-level inference.

Governance-wise, it describes the governance structure that goes with augmentation: Selective delegation, Expert verification, Shared integration routines, Explicit accountability, Protected learning opportunities, and continuity of future capacity to critically review.

2. Literature Review

2.1 From task acceleration to collective performance

GenAI productivity research consistently reports substantial gains on bounded knowledge tasks, yet the reported gains differ by baseline skill, task structure, and output criterion. The highest level of consensus is about acceleration generation decreases first draft time, search effort and the formulation of routines (Doshi & Hauser, 2024; Noy & Zhang, 2023). Quality is less agreeable if it is of a multi-dimensional nature, or if work needs to be coordinated afterwards. Completion time is not the only event that is not accounted for in collective performance, as there are also handoffs, mutual monitoring, revision ownership, and contributions by specialists who are integrated into the performance. Hybrid-intelligence scholarship, therefore, posits value as better joint performance from the careful and intentional synthesis of human and machine capabilities (Dellermann et al., 2019). Communication, role clarity, and shared mental models are also needed coordination resources added to the list of necessary resources for machine-teammate research (Seeber et al., 2020).

The level of analysis is the source of the fact that there is a contradiction between individual gains and organizational value. A user benefits from receiving the local time; the team takes on exception-handling and verification and standardization expenses. Trust research indicates that the correct level of trust, rather than excessive trust, is the key to performance (Glikson & Woolley, 2020). Digital systems also re-distribute control, visibility, and contestability among work groups, as shown by algorithmic management research (Kellogg et al., 2020). A faster individual task only creates collective value if there is less rework downstream, shared knowledge is available, and accountability is still legible.

This difference in level is also responsible for the differential outcome measures. Novelty, factual accuracy, maintainability and collective diversity have different standards, whereas output quantity favors generation. A team that generates more alternatives that narrows the solution space is able to record throughput without increasing their

solution space. Thus, a quality-adjusted and a coordination-adjusted denominator is needed for productivity.

2.2 Expertise as a complementary production resource

Expertise complementarity refers to performance improvement produced by nonredundant strengths and partially uncorrelated errors. Contextual diagnosis and causal reasoning, normative judgment, and the identification of boundary cases are provided by domain experts (Hemmer et al., 2025; Inkpen et al., 2023). GenAI provides quick generation, syntactical variation, retrieval like association and scalable iteration. Model relationship is configurational; same model helps with one task and hinders with another. Empirical evidence on joint classification indicates that delegating based on a decision algorithm is more effective than delegating based on humans' relative advantage, when humans misjudge their relative advantage (Fügener et al., 2022). AI advice has been shown to increase average accuracy in repeated situations, but lead to a diminished number of unique human pieces of knowledge, creating a trade-off between conforming to the AI and maintaining informational diversity within a community (Fügener et al., 2021).

Prompts and validation are also driven by expertise. In the case of high performers, the baseline is already high, so they are not likely to gain much on routine outcomes; however, their knowledge aids in selecting, correcting and integrating, which are more reliable, outcomes (Brynjolfsson et al., 2025; Inkpen et al., 2023). Low-expertise users receive templates and procedural scaffolds, and experience more fluent errors. This forms a dual-sided system; GenAI helps close execution gaps and expert oversight ensures quality of outputs. Likewise, decision-structure research indicates that allocation of authority is contingent on the decomposability of a task, the explainability of the task, and the relative specificity of humans' knowledge (Shrestha et al., 2019). Knowledge is still useful if the workflow is structured to allow the experts some room for diagnosis and review, rather than ceremonial approval.

2.3 Delegation, trust calibration, and task–technology fit

Delegation is the operational link between complementary resources and performance. Agentic information systems receive tasks and produce outputs, and in some cases take up tasks and reassign them (Baird & Maruping, 2021). To be productive, it is necessary to know how well the system can function, to be aware of its own limitations, and to have a path back down from the escalation. Systematic synthesis reveals that the aversive nature of the task changes as a function of task framing, exposure to errors, explanation, and decision control that is retained (Burton et al., 2020). Algorithmic appreciation and long-term aversion following a visible error have been reported in behavioral studies (Dietvorst et al., 2018; Logg et al., 2019). In subjective tasks, individuals are more prone to resist algorithms, while in objective and measurable tasks, individuals are more prone to accept algorithms (Castelo et al., 2019).

Calibration is made challenging by the fact that there are many different prompts and instances of GenAI, resulting in varying levels of output quality. The verbal cue 'confidence' is not a good indicator of accuracy. Thus, the classification of risks, the checks of origin and the validation standards associated with the consequences of tasks are an essential part of the process of delegation decisions (Lai et al., 2025). There is a trade-off between more human control over the process of an AI's work and less in more complex tasks and more accountability over the work by the external world and less by the human for more well-specified tasks. This task–technology fit perspective is based on the premise of not adopting universal adoption targets, but rather conditional adoption rules.

2.4 Sociotechnical integration and quality preservation

Team productivity emerges in a sociotechnical system of roles, tools, routines, incentives, and team knowledge flows. Work Design research indicates that digital technology concurrently alters autonomy, feedback, skill variety, relational contact, and workload (Parker & Grote, 2022). AI can be useful when integrating technical aspects with human skills, coordination, and learning routines within the organizations (Mikalef & Gupta, 2021). At the individual level, augmentation is more positive in the context where AI use increases task resources (as compared to AI decreasing them), and when individuals have discretion in decisions that follow from AI use. This research points away from access to the model towards organizational structure.

There are four integration routines for quality preservation. First, teams define the points in the workflow where AI is used and its exit. Second, the outputs generated are owned and are up for review. Third, reviewers are not concerned with surface plausibility. Fourth, teams transform miscorrected errors into team prompts, tests, and playbooks. These routines disappear when they're not in place, and there are expert bottlenecks. Too much automation also takes the learning opportunities from junior employees, thus compromising the pool of reviewers in the future. Capability renewal, in turn, is a part of sustainable augmentation of productivity.

2.5 Unresolved gap and analytical propositions

While these areas are well documented in the literature, the links between them are not well established at team level. Experiments isolate tasks; Surveys measure attitudes; Tensions are outlined in organizational theories but not there to be diagnosed in an operation. Very few studies differentiate a productivity effect from a collaboration effect and even fewer studies look at if a high adoption rate can exist alongside low trust and purposeful exclusion of significant parts of the workflow.

There are three analytical propositions which follow. Firstly, adoption will outpace favourable sentiment as

availability and use will outpace reliable assimilation. Second, reported time and productivity gains will be greater than collaborative improvement since there have to be extra routines involved in coordinating. Third, knowledge will be expressed as measured skepticism and as a selective rejection of tasks and not as a general rejection of knowledge. The propositions serve a secondary analysis across the sections that will be carried out as a cross-sectional study to look for patterns of alignment, rather than causal treatment effects.

3. Conceptual and Theoretical Framework

Human-capital complementarity and task–technology fit explain the production conditions under which GenAI supports rather than displaces professional expertise. The performance criterion is complementarity, which occurs when the human and model capabilities are nonredundant, the error structures are different and there is still expertise available to identify exceptions (Hemmer et al., 2025). Task–technology fit is the specification of the allocation rule. GenAI can help generate new ideas, creating variations, analyzing a routine, and executing a task with some restrictions, but it is not helpful in situations where there is uncertainty, or where the task is a consequence of the input. Task Decomposition, Prompt Specification, Boundary Recognition, and Escalation Judgment are task contributions of Domain Experts. The framework thus does not consider expertise as a demographic control or something that is generally moderated but is rather regarded as an active routing resource (Cantini & De Mauro, 2026).

The locally useful output turns into a collective contribution through the use of socially shared diagnosis and mode of action, as described by sociotechnical systems reasoning and calibrated-reliance theory. The material created goes into an organizational structure of roles, points of review, storage, handoff, documentation, and accountability (Parker & Grote, 2022). Verification checks the factual accuracy and contextual appropriateness, and reintegration brings corrections and provenance to the attention of colleagues. Trust then does not consist of blind faith, but of generalised rejection. It's a threshold that connects perceived models ability to the consequence of the task. Complementary value is created by experienced professionals when they are doubted and thus subject to critical attention of consequential products, while when they are not doubted, useful experimentation is stifled, and when they are trusted, plausible errors are not rejected. Collaboration, burden of review, and quality retention are therefore separate consequences and not synonyms of adoption (Górriz et al., 2023).

The integrated Expertise Delegation Integration model starts from four theoretical lenses and then links the expertise and task conditions to the AI delegation of tasks, the verification and reintegration of tasks, and the team's outcomes. The desired outcome bundle is faster flow, lower coordination costs, better collaboration, and maintained quality. Learning feedback provides feedback on corrections and improved boundaries, and provides review rationales back to the expertise base to rejuvenate the expertise that will serve as a basis for subsequent delegation. The empirical context is provided by cross-sectional secondary developer data, indicators that are specific to the denominator, and indicators of organizational governance. This structure is shown in Figure 1 as an interpretative conceptual model instead of an estimated causal model because the available aggregates cannot be used for inference about individual users, tasks, teams and outcomes, and are not used to estimate the pathways. Its analytical purpose is pattern matching: When it comes to adoption, it is pattern matching to identify the one defined by rising adoption and high reliability friction with weak collaboration; when it comes to reintegration, it is pattern matching to identify the one defined by shared reintegration; and when it comes to learning feedback, it is pattern matching to identify the one defined by expert verification and delegation selectiveness.

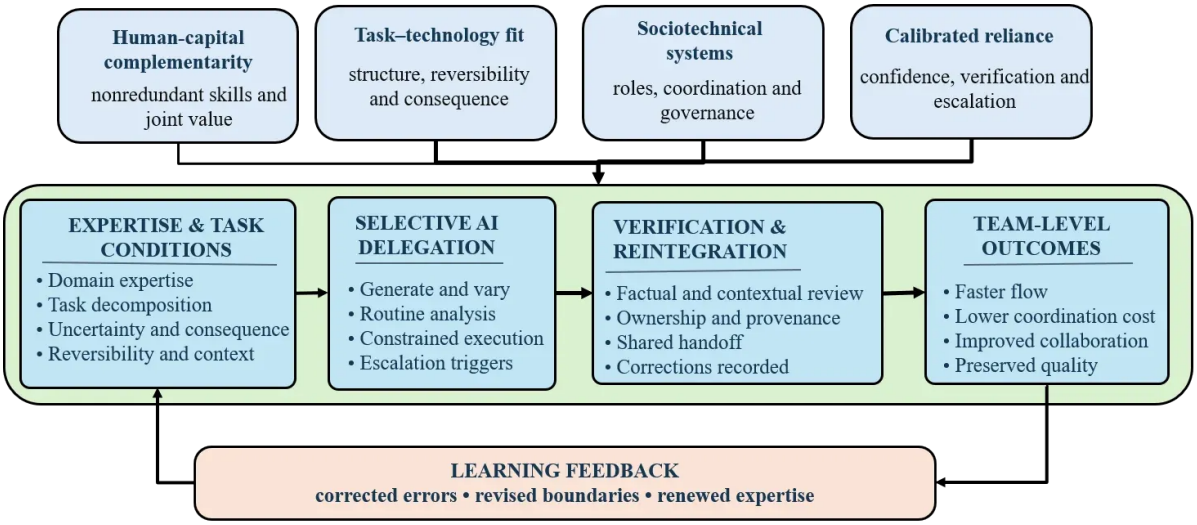
4. Methodology

4.1 Research design

The research was conducted using secondary data, cross-sectional and quantitative design. The publicly available aggregated results of the 2025 SO Developer Survey were reorganized in this way by the EDI framework. The design was chosen as it captures the research problem related to alignment of adoption, expertise, delegation, and team outcomes during a time of broad diffusion of GenAI. The reported response distribution was for a specified survey item or subcategory, rather than for a single respondent or team, employer or country. Depending on this, all relationships are descriptive and ecological.

4.2 Study area, population, and analytical sample

The study area consisted of the global software-development sector, which was represented in the 2025 Stack Overflow Developer Survey. This number 62 indicates geographical and platform coverage, and the number of questions in the source survey was over 49,000 respondents from 177 countries. The target audience consisted of digitally-savvy developers and tech people who are part of the developer ecosystem globally. The analytical sample is based on an AI workflow-stage module, which was completed by 11,202 respondents, since this module directly observed delegation boundaries. Fewer than half of the published modules had supplementary indicators, which only provided productivity and agent-impact, frustration, concern, and trust indicators. Hundreds of respondents completed the items, and this volume of responses was used to inform the analysis in some instances; in others, items were routed or not responded to.



4

Fig. 1. Expertise Delegation Integration framework

4.3 Variable construction and measurement logic

Percentages listed on the survey dashboard were used to create the variables. Current or planned use was considered adoption. There were positive attitudes towards AI tools. Individual performance has been seen to have improved task time and increased productivity. Team integration: This was proxied by the enhanced collaboration between agent users. Expertise was symbolized by groups of years of professional experience and extremes of experience of trust that were published. Delegation fit used was not used at deployment and monitoring, was not used at project planning, and perceived difficult with complex tasks. Quality risk was a mix of accuracy, privacy, near-correct output and debugging-burden indicators. The constructs are operationalized in Table 1 because there are three distinct parts of the framework: mechanisms, outcomes, and guardrails. Here, percentage-point difference instead of percentage change was the main comparison due to all indicators being proportions. The adoption–sentiment gap was the difference between 2025 adoption and 2025 positive sentiment. The gap in productivity between agents and users was the agent-user productivity percentage minus the collaboration percentage. These amounts represent distance between constructs and not necessarily that the same respondents provided both responses.

Table 1. Operational definitions and decision rules

Construct	Indicator	Scale	Decision rule
Adoption	Uses or plans AI tools	Percentage	Higher value = wider diffusion
Assimilation	Positive sentiment	Percentage	Compared with adoption, not treated as output
Individual productivity	Favorable productivity; time reduction	Percentage	Higher value = stronger local benefit
Team integration	Improved collaboration	Percentage	Compared with individual benefit
Trust calibration	Trust, distrust, extreme trust by experience	Percentage	Skepticism interpreted with task exclusions
Delegation fit	workflow non-use; complex-task struggle	Percentage	Higher exclusion = narrower AI boundary
Quality guardrail	accuracy, privacy, repair burden	Percentage	Higher value = greater governance need

4.4 Analytical procedures

The analyses were carried out in 4 steps. First, percentages, labels, and bases for each item were converted to a structured workbook and compared to the survey modules that were shown. Secondly, trend comparisons were made descriptively for adoption and sentiment for 2023–2025. Third, percentage-point differences were used to compare individual benefits and collaboration, and trust and distrust. Fourth, pattern matching was used to match the direction of each indicator set with the EDI propositions.

Analytical objective was followed by chart form. Annual directions were indicated by lines and the benefits and constraints were indicated by horizontal bars and the trust categories were contrasted by clustered bars. Zero was used as a starting point for all percentage bars and the axes of all line

charts were maintained at zero to avoid creating an exaggerated effect. The values were plotted directly, no smoothing was applied, nor was any interpolation. An overall synthesis was conducted to categorize each of the analytical objectives as to whether it was aligned or not aligned with the EDI mechanism.

There was no respondent level regression or significance test performed, nor was there a mediation model or causal estimator applied, due to the lack of joint distributions and sampling weights at the respondent level. The difference in the arithmetic contrasts do not refer to any kind of descriptive differences. No imputations were made for missing data. Categorization in this document does not yield synthetic scales for categories that have also not been combined in this manner because the questions are not worded the same or the responses are based on different types of questions.

4.5 Quality control, validity, and bias management

Dual entry verification of each plotted value, explicit documentation of response base and formula checks for quantity percentage-point gaps, and one source data table per graph were used for data quality control. Each indicator was linked to a pre-established element of the framework to further enhance the construct validity. In the interest of clarity in the discrimination, the distinction between adoption, sentiment, productivity, and collaboration was kept separate. Avoiding claims about the behavior of individuals from aggregate distributions helped to manage ecological bias. Bias of self-selection, self-reporting, recall, platform composition and question-routing have not been ruled out. There are also associations among perception items that are affected by common-method bias.

Checks of robustness included denominator sensitivity and interpretation-reversal checks. The narrowest eligible population of each module was used to re-read each conclusion. If the direction of a finding was not affected by that restriction, then the finding was retained. Plausible alternative explanations such as novelty effects, heterogeneity in tools and occupational composition were also investigated. None of these checks performed any transformations of descriptive patterns into causal estimates.

4.6 Ethics and reproducibility

All of the secondary data analysed were anonymised and publicly reported, and there was no contact with the participants, any identifying information or any re-identification was attempted. Primary collection was under the control of the public reporting and privacy controls in the source survey. Reproducibility is achieved by an accompanying workbook that includes each value, denominator, formula, chart and source URL. Interpretations continue to be confined to the community of developers that they represent and the language used in the published material.

5. Results

5.1 Coverage and expertise composition

Coverage is the response base that is associated with each indicator, and helps to ensure that no module or indicator is counted by all 49,000+ international respondents. This is important since over-stating the support for the delegation and quality findings would occur with denominator inflation. Table 2 illustrates the 11,202-person workflow sample as the foundation for analysis, and indicates the larger ancillary modules. The workflow module is the smallest, but most direct measure of delegation. The top complementary module is productivity (+ 31,636), followed by frustrations (+ 31,476), while 49,000+ is source reach only. The hierarchy widens the interpretation but doesn't add together non-equivalent denominators.

Table 2. Data coverage, analytical response bases

Analytical module	Reported response base	Analytical role
Overall survey	49,000+	Source coverage only; not the analytic sample
AI workflow stages	11,202	Core analytical sample for delegation boundaries
AI-agent impacts	12,823	Supplementary time, productivity, and collaboration indicators
AI productivity perceptions	31,636	Supplementary broader productivity indicator
AI frustrations	31,476	Supplementary verification-burden indicators
AI concerns	28,930	Supplementary quality and governance-risk indicators
Trust by experience	25,093 across three groups	Supplementary expertise-calibration profile

Experience composition gives a way to compare the profiles in the context of judging whether or not it is professionally appropriate for them to rely on each other. The number of modules is not used to determine workforce shares or individual transitions. Table 3 shows comparisons between the early career, mid-career, and experienced groups on the use, sentiment, trust, and complex-task items. The largest and smallest groups are both lower-experience respondents, larger than they are twice, while mid-career respondents are smaller. There are only 102, 75, and 100 fewer counts in the three groups, respectively, for those using AI with texts and those using AI to complete more complex tasks. This stability is important for the description of the adjacent module, but the aggregates do not describe repeated individual respondents. The profile is descriptive, not a workforce distribution.

Table 3. Experience-group response profile across AI modules

Experience group	AI use	Sentiment	Trust	Complex tasks
Early career, 1–5 years	6,360	6,293	6,254	6,258
Mid-career, 5–10 years	5,997	5,957	5,931	5,922
Experienced, 10+ years	13,001	12,941	12,908	12,901

5.2 Adoption–sentiment divergence

Adoption–sentiment divergence asks the question of whether there is a correlation between the spread of adoption and an overall more positive experience. Growth sets the groundwork, but does not create productive assimilation. Use or planned use shown in Figure 2 is from 2023 to 2025. Adoption rose from its lowest value of 70% in 2023 to 76% in 2024 and its highest value of 84% in 2025. The growth jumped by 6 to 8 per cent despite uncertainties over the operation. The case of an upward trend is thus indicative of ongoing diffusion.

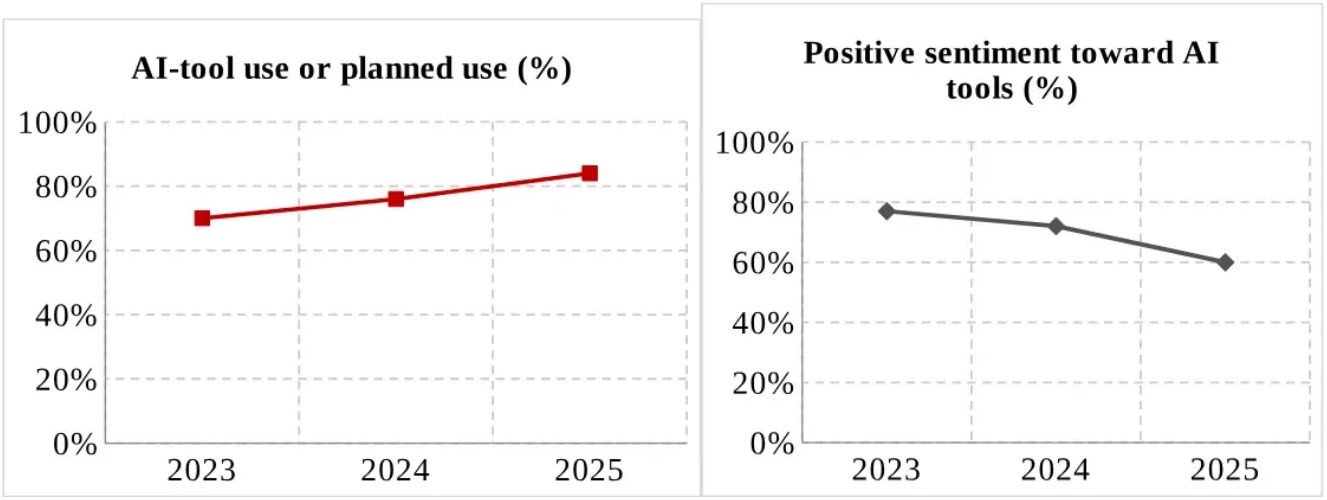


Fig. 2. AI-tool use or planned use, and Positive sentiment toward AI tools, 2023–2025

Sentiment is added to the assimilation assessment to determine if experience increased when adoption increased. In addition, there's a comparison of favorable sentiment over the same period in Figure 2. In the last three years, the levels of sentiment have decreased from 77% in 2023 to 72% in 2024 and 60% in 2025. The decline between adoption and sentiment was all of ten points to twelve points, resulting in an adoption–sentiment gap of 24 points in 2025. This was precisely when reliability and flow costs had become more critical, so increased use came at the time of increased critical scrutiny of reliability and flow costs. The change is the most evident assimilation deficit.

5.3 Individual productivity and team conversion

This leads to the conversion dimension, which relates the performance of the local effort to the performance of the collaborative effort, and recognizes that team productivity demands coordination gains. Benefits are shown for both the agent-impact modules and the productivity modules in Figure 3. The fewest hours are spent at 70% (reduced), followed by increased productivity at 69%, favorable broader productivity at 52%, and collaboration at 17% (lowest). The one-point time productivity difference is a measure of direct local benefit, while the 52-point productivity collaboration gap is a measure of the primary conversion bottleneck. While speed is common, shared benefit is still not widespread and requires review, ownership, documentation, and handoff processes. The gap directly aligns to the objective related to team conversion.

Adopting and delegating workflow responsibility are two different things in the adoption ecosystem. The profile is shown in Table 4. The numbers for out-of-the-box use are highest for ChatGPT at 82%, and GitHub Copilot at 68%. On the other hand, 52% reported that they do not have any plan to use agents or are choosing simpler tools. However, 38% said they didn't have a plan to use agents, and 52% reported no agent use or that they prefer simpler tools. Even at such a low value, there is quite a bit of resistance to autonomous

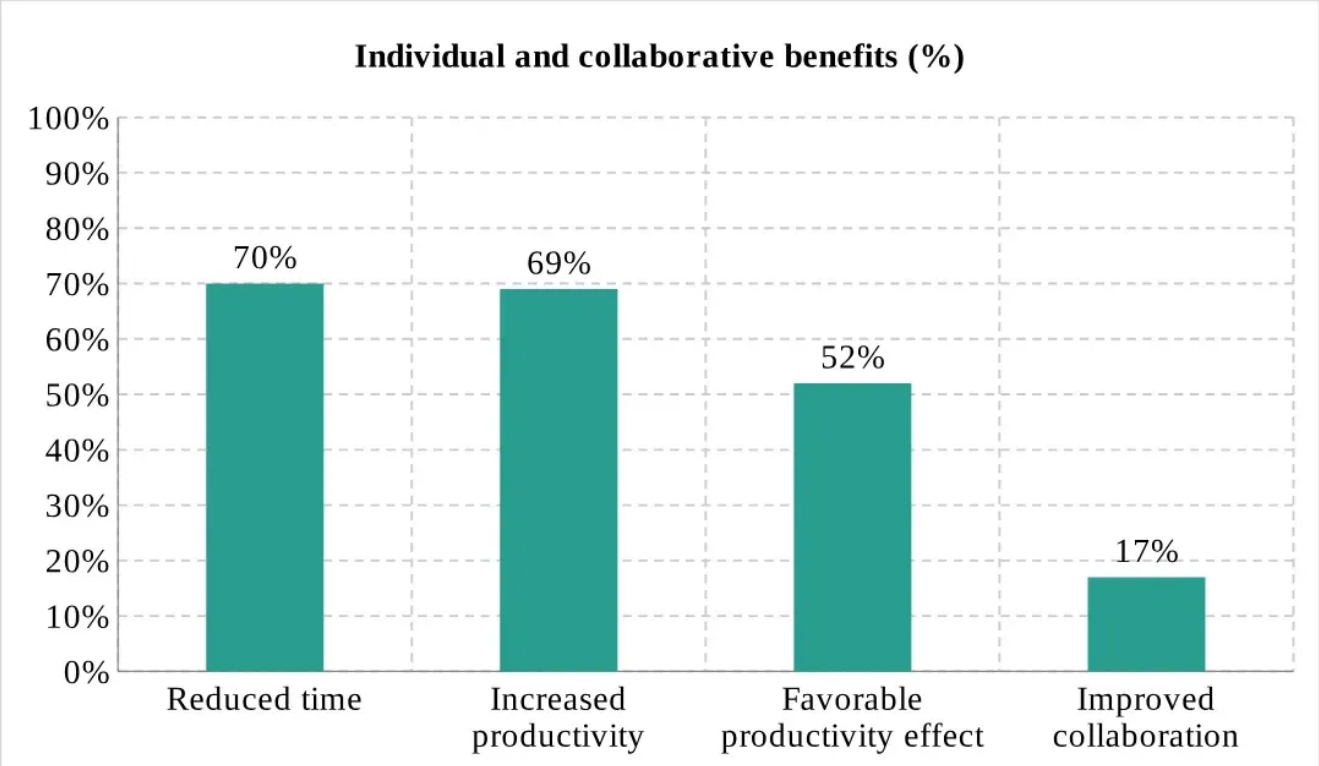


Fig. 3. Reported individual and collaborative benefits

orchestration. Thus, interaction between general-assist and agent is more extensive than sustained agentic delegation and mature team augmentation. The uptake of tools then exaggerates the extent of workflow integration.

Table 4. Agent and assistant adoption ecosystem

Indicator	Percentage	Interpretation
ChatGPT used as an out-of-the-box tool	82	Dominant general assistant
GitHub Copilot used as an out-of-the-box tool	68	Strong coding-workflow presence
No agent use or preference for simpler tools	52	Agentic integration remains limited
No plan to use AI agents	38	Substantial resistance to autonomous workflows

5.4 Expertise-calibrated trust

Trust calibration is a measure of whether trust is associated with the awareness of model limits. Skepticism produces complementary value if it guides verification to consequential outputs because of expertise. Overall trust is compared to the extremes of the groups of experience in Figure 4. By thirteen points, Distrust outstrips Trust, at 46% (up 13 points). Only 3.1% is high trust, the lowest being among experienced respondents (2.6%), while experienced high distrust is 20% overall. The most striking difference would be within the experienced group and would thus favor skeptical reliance without individual verification behavior. Low confidence does not mean that it is rejected altogether.

5.5 Delegation boundaries

Delegation boundaries are a way to find work under human control and where consequence and system fit limit GenAI usage. Figure 5 shows a relationship among workflow exclusions and complexity of the tasks. The percentages for deployment and monitoring non-use are the highest (76%), and project planning is the next highest (69%), and complex-task struggle is the lowest (29%). That 47-point difference means that delegation, as well as complexity, is impacted more by accountability and operational consequence. The strongest wall is around stages of errors going into live systems. The second most robust boundary is the project planning.

5.6 Reliability friction and quality preservation

Reliability friction refers to unexposed work related to uncertainty in outputs and governance risks. Verification decreases overall productivity and is an increasing burden on review for knowledgeable staff. The accuracy, privacy, near-correctness, and debugging concerns are compared in Figure 6. The most accurate at 87% is followed by privacy or security at 81%. Nearly right or extra debugging time are the two most common almost right solutions - 66% and 45% respectively. There is, however, a cost of debugging, that is, labour required and the speed lost. The six-point difference between accuracy and privacy is a measure of two widespread governance pressures, while the 42-point difference between accuracy and experienced labour cost distinguishes general concern from experienced labour cost. There is still an area of near-correct output, which is a verifiable problem.

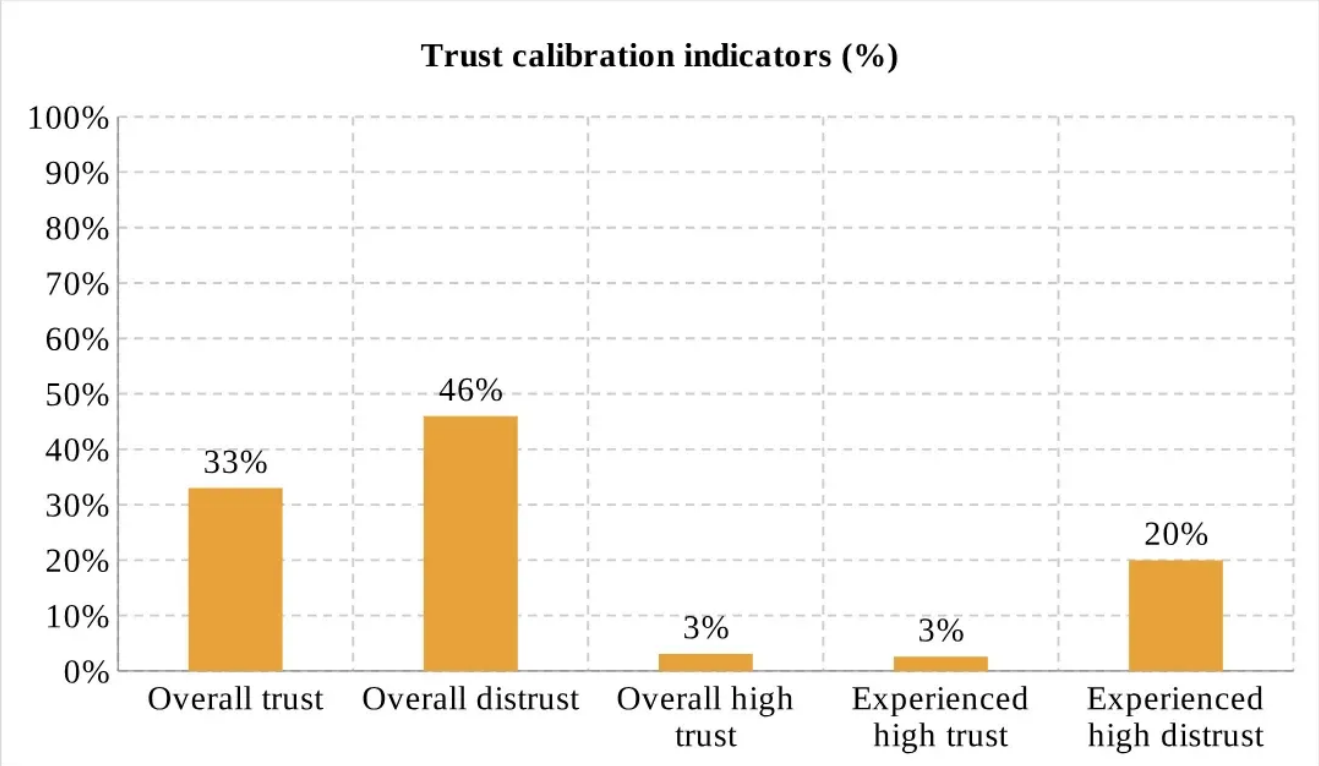


Fig. 4. Trust calibration indicators

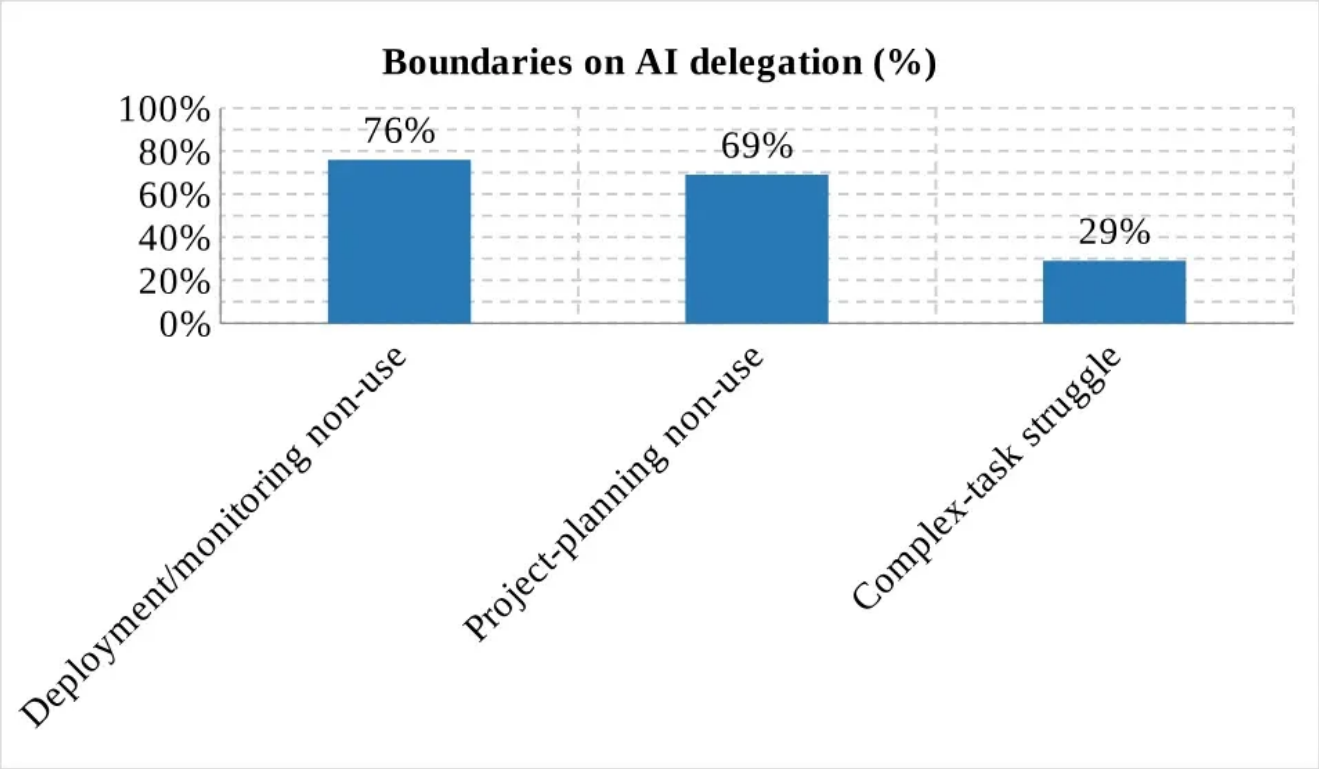
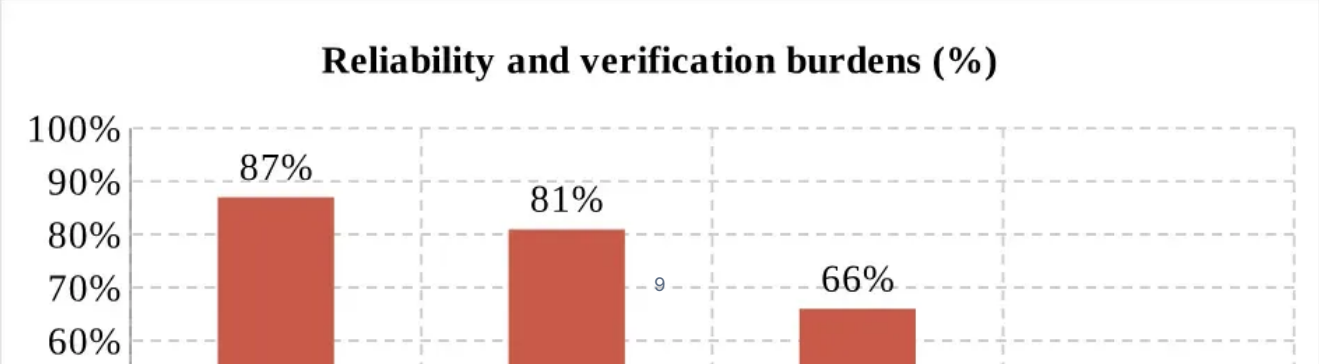


Fig. 5. Boundaries on AI delegation



Pattern matching combines the dimensions and differentiates between augmentation and automation. The results of the directionality tests are summarized in Table 5. When diffusion goes up as sentiment goes down, adoption prevails over assimilation; when individual benefits go up while collaboration goes down, integration is the lowest level; and when high trust is low on experienced, and high exclusions at consequential, the level of risk sensitivity is high - this is risk-sensitive delegation. The setup is transitional. The use of Selective and Emerging Expert Control is strong, and the use of systematic reintegration within shared team processes is weak. Expert control is still vital with regard to quality preservation.

Table 5. EDI pattern-matching assessment

Analytical objective	Observed directional pattern	EDI assessment
Adoption and assimilation	Diffusion rises while sentiment falls	Adoption exceeds assimilation
Individual-to-team conversion	Individual benefits exceed collaboration	Integration is the principal bottleneck
Expertise calibration	Experienced respondents report very low high trust and marked high distrust	Expertise aligns with skeptical reliance
Delegation and quality boundaries	Consequential workflow stages face the strongest exclusion amid high reliability concern	Delegation is selective and risk-sensitive

6. Discussion

The results suggest that the challenge in the organization is no longer getting access to GenAI but rather creating an architecture for converting. When sentiment is declining and adoption is picking up, it's a sign of a technology getting into routine exposure. Users at the same time are presented with benefits and failure modes. The execution measures have built-in trust due to controlled studies reporting higher completion rates and better quality, and field studies reporting higher gains for less experienced workers. The present analysis introduces a distinction at the level of the location: commonality at the local level is limited at best, but collaborative improvement at the local level is far less common.

The EDI framework's claim that integration is a mediator in the relationship between team value and the productivity-collaboration gap is supported by the 52-point productivity-collaboration gap. Team agreement issues of ownership, review thresholds, documentation, and escalation remain as with any individual task. If the routines are not in place, saved time comes back as checking, repair, and coordination. The outcome is an extension of the automation-augmentation paradox where the practical conflict between the output and team levels is found. Augmentation is not a characteristic of the model or prompt. It is a collective achievement created after generation.

This is confirmed by the results of trust. There is more mistrust than trust overall, and there is very low trust in highly experienced respondents. This is not to be confused with technology rejection. As adoption increases, confidence decreases, indicating differentiated reliance - use the system, check results, and outside of further subsequent processes. This pattern is consistent with the complementarity studies that indicate that the performance of a joint is dependent on the knowledge of the expert, the tuning of the model and the allocation rules (Hemmer et al., 2025). It also serves to balance optimism of the experimental scientist with frustration in the workplace. Note that while GenAI is quite effective within its bounds of a well-defined task frontier, near-miss answers add up to expensive ambiguity outside the task frontier.

The exclusions of the tasks indicate how professionals draw the frontier. The resistance is more towards deployment and monitoring than towards the notion that AI is not adept at complex tasks. Consequence and accountability seem to be more important factors than cognitive challenge alone. Project planning is also secure, demonstrating the importance of prioritizing within context and negotiating amongst stakeholders. These boundaries are those of a sociotechnical account that offers, in addition to correction, proper accountability. Moreover, they fit the interaction irony situations reported, which verify that repair takes up the time that automation saves (Simkute et al., 2024).

The theoretical suggestion is that there are a production function and a renewal function in the expertise complementarity. Experts enhance existing outputs by breaking them down and reviewing them, and junior involvement adds capacity for the future to undertake those activities. Short-run fragility for novices is achieved by removing developmental tasks; long-run fragility is achieved. The implication for management is even more straightforward: adoption measures need to have other measures to accompany them, such as review burden, escaped defects, collaboration, and learning. A team that uses technology, but is not an augmented team.

The pattern-matching design illustrates the usefulness and the boundaries of public aggregate survey results. Convergent indicators show a consistent configuration, but not individual pathways and effects on the firm level. Self-report outcomes also differ from audited output, quality, and cycle-time records. Future tests will need to include team rosters, expertise measures, review traces, linked task logs and longitudinal quality outcomes. The present results provide a baseline diagnostic measurement, but do not provide an estimate of causality.

Presented below are several implications for theory. First, expertise is not just a moderator of a technology–performance relationship, but also a resource for routing and controlling a relationship. Second, collaboration is not the same as generation, it's a different stage of production that comes after generation. Thirdly, mistrust has conditionally valuable uses in that it draws attention to the consequences of what is being done without stifling beneficial experimentation. These

propositions extend the concept of complementarity to the beyond static comparison between the accuracy of humans and models.

For use, the results contradict enterprise-wide usage quotas as key transformation measure. A scorecard that is meaningful is coupled with adoption and defect escape, review time, handoff latency, collaboration and skill building. It also logs any work done to verify the record. Highly focused review by the senior staff causes queues and takes away valuable knowledge from design, mentoring and exception resolution. These are distributed protocols, automated tests and explicit escalation thresholds, which lessen that concentration without losing accountability of ownership.

7. Conclusion

The study focused on the question of why GenAI is adopted widely, but yields disparate team-level productivity. It aimed at providing an explanation of the conversion process based on the concepts of expertise complementarity, selective task delegation, trust calibration and sociotechnical integration. A secondary analysis of published 2025 developer-survey data was conducted to provide a cross-sectional analysis in a framework that was similar to the EDI pattern-matching framework but without item-specific denominators and without respondent-level causal claims.

A transitional configuration is revealed by the results. Eighth use/planned use at 84% and positive sentiment at 60%. The benefits per individual were significant: 70% of agent users said that using agents reduced the time they spend, 69% said it improved productivity. Only 17% indicated that the collaboration of the team was improved. Distrust outstripped trust with experienced developers reporting limited levels of high trust, while respondents established clear boundaries for deployment, monitoring, and project planning. Net quality is threatened by unrestricted automation by accuracy, privacy, plausible error, and debugging burden.

The contribution is a multi-level description of augmentation. Knowledge was transformed into output to enable team productivity, including setting task boundaries, monitoring exceptions, reconnecting outputs, and refreshing shared knowledge. Self-selection, self-reporting, unequal module bases, developer-community concentration, and aggregate data that cannot be used to estimate causality are limitations. Longitudinal team records, objective measures of quality, task-level exposures, reviewer workload, and quasi-experimental designs of rollout are needed in future research. Such work would be a test of the durability of individual speed and of the sustainability of the expertise that augmentation relies upon when it turns into durable collective performance and whether protected learning opportunities maintain this expertise.

The real-world takeaway is that, while GenAI supports in team systems, it is not a self-contained alternative to such systems in a variety of contemporary and consequential knowledge-work contexts, but instead, it remains a governed contributor in such team systems.

8. Recommendations

They need a task-allocation register which categorizes GenAI usage based on reversibility, consequence, uncertainty and expertise. Low-risk generation should be part of assisted flows and deployment; monitoring, strategic planning, and accountability decisions for external parties should have named human owners and explicit approval gates. Provenance and review status, along with material corrections, are necessary for each AI-assisted artifact.

There is a need for separate metrics on time saved, rework, escaped defects, reviewer load, collaboration, and knowledge transfer for performance dashboards. An increase in usage in any of these measures indicates a pressure of automation, not augmentation. Have common prompt libraries, test suites, exception logs, and escalation paths to ensure that corrections become organizational knowledge.

The protection of expert capacity is also a requirement. Junior staff should receive supervised diagnostic and review work rather than their complete absence from the foundation work. The experienced reviewers require workload limits and recognition for verifying activity. Quarterly workflow audits should review benefits reported with quality outcomes and adjust boundaries of delegation as necessary. These actions directly address the observed productivity collaboration gap, low trust, strong task exclusions, and high reliability concern.

Implementation should include a named person, a baseline of 30 days, and a review cycle of 1 month per quarter. Delegation expansion can only be approved when there is a concurrent improvement of quality, coordination, and learning.

Author Contributions

Conceptualization, methodology, data curation, formal analysis, visualization, writing original draft, writing review and editing: Anil Doshi and Devika Narayan. Both authors approved the final manuscript.

Conflict of Interest

The authors declare no conflict of interest.

References

Alavi, M., Leidner, D. E., & Mousavi, R. (2024). Knowledge Management Perspective of Generative Artificial Intelligence. Journal of the Association for Information Systems, 25(1), 1–12. https://doi.org/10.17705/1jais.00859

Baird, A., & Maruping, L. M. (2021). The Next Generation of Research on IS Use: A Theoretical Framework of Delegation to and from Agentic IS Artifacts. MIS Quarterly, 45(1), 315–341. https://doi.org/10.25300/MISQ/2021/15882

Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). Managing Artificial Intelligence. MIS Quarterly, 45(3), 1433–1450. https://doi.org/10.25300/MISQ/2021/16274

Boussioux, L., Lane, J. N., Zhang, M., Jacimovic, V., & Lakhani, K. R. (2024). The Crowdless Future? Generative AI and Creative Problem-Solving. Organization Science, 35(5), 1589–1607. https://doi.org/10.1287/orsc.2023.18430

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

Burton, J. W., Stein, M., & Jensen, T. B. (2020). A systematic review of algorithm aversion in augmented decision making. Journal of Behavioral Decision Making, 33(2), 220–239. https://doi.org/10.1002/bdm.2155

Cantini, A., & De Mauro, A. (2026). Better Prompts, Better Usefulness: A Systematic Review and Experimental Evaluation of Structured Prompting Techniques in Large Language Models. Big Data and Cognitive Computing, 10(7), 224. https://doi.org/10.3390/bdcc10070224

Castelo, N., Bos, M. W., & Lehmann, D. R. (2019). Task-Dependent Algorithm Aversion. Journal of Marketing Research, 56(5), 809–825. https://doi.org/10.1177/0022243719851788

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

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

Doshi, A. R., & Hauser, O. P. (2024). Generative AI enhances individual creativity but reduces the collective diversity of novel content. Science Advances, 10(28), eadn5290. https://doi.org/10.1126/sciadv.adn5290

Faraj, S., Pachidi, S., & Sayegh, K. (2018). Working and organizing in the age of the learning algorithm. Information and Organization, 28(1), 62–70. https://doi.org/10.1016/j.infoandorg.2018.02.005

Fügener, A., Grahl, J., Gupta, A., & Ketter, W. (2021). Will Humans-in-the-Loop Become Borgs? Merits and Pitfalls of Working with AI. MIS Quarterly, 45(3), 1527–1556. https://doi.org/10.25300/MISQ/2021/16553

Fügener, A., Grahl, J., Gupta, A., & Ketter, W. (2022). Cognitive Challenges in Human–Artificial Intelligence Collaboration: Investigating the Path Toward Productive Delegation. Information Systems Research, 33(2), 678–696. https://doi.org/10.1287/isre.2021.1079

Glikson, E., & Woolley, A. W. (2020). Human Trust in Artificial Intelligence: Review of Empirical Research. Academy of Management Annals, 14(2), 627–660. https://doi.org/10.5465/annals.2018.0057

Górriz, J. M., Álvarez-Illán, I., Álvarez-Marquina, A., Arco, J. E., Atzmueller, M., Ballarini, F., Barakova, E., Bologna, G., Bonomini, P., Castellanos-Dominguez, G., Castillo-Barnes, D., Cho, S. B., Contreras, R., Cuadra, J. M., Domínguez, E., Domínguez-Mateos, F., Duro, R. J., Elizondo, D., Fernández-Caballero, A., … Ferrández-Vicente, J. M. (2023). Computational approaches to Explainable Artificial Intelligence: Advances in theory, applications and trends. Information Fusion, 100, 101945. https://doi.org/10.1016/j.inffus.2023.101945

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

Hillebrand, L., Raisch, S., & Schad, J. (2025). Managing with Artificial Intelligence: An Integrative Framework. Academy of Management Annals, 19(1), 343–375. https://doi.org/10.5465/annals.2022.0072

Inkpen, K., Chappidi, S., Mallari, K., Nushi, B., Ramesh, D., Michelucci, P., Mandava, V., Vepřek, L. H., & Quinn, G. (2023). Advancing Human-AI Complementarity: The Impact of User Expertise and Algorithmic Tuning on Joint Decision Making. ACM Transactions on Computer-Human Interaction, 30(5), 1–29. https://doi.org/10.1145/3534561

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

Kellogg, K. C., Valentine, M. A., & Christin, A. (2020). Algorithms at Work: The New Contested Terrain of Control. Academy of Management Annals, 14(1), 366–410. https://doi.org/10.5465/annals.2018.0174 Lai, C.-H., Chen, Y.-J., & Chen, Z.-P. (2025). Fine-Tuning Generative AI with Domain Question Banks: Evaluating Multi-Type Question Generation and Grading. Applied Sciences, 15(24), 13050. https://doi.org/10.3390/app152413050

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

Mikalef, P., & Gupta, M. (2021). Artificial intelligence capability: Conceptualization, measurement calibration, and empirical study on its impact on organizational creativity and firm performance. Information & Management, 58(3), 103434. https://doi.org/10.1016/j.im.2021.103434

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

Parker, S. K., & Grote, G. (2022). Automation, Algorithms, and Beyond: Why Work Design Matters More Than Ever in a Digital World. Applied Psychology, 71(4), 1171–1204. https://doi.org/10.1111/apps.12241

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

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

Shrestha, Y. R., Ben-Menahem, S. M., & von Krogh, G. (2019). Organizational Decision-Making Structures in the Age of Artificial Intelligence. California Management Review, 61(4), 66–83. https://doi.org/10.1177/0008125619862257

Simkute, A., Tankelevitch, L., Kewenig, V., Scott, A. E., Sellen, A., & Rintel, S. (2024). Ironies of Generative AI: Understanding and Mitigating Productivity Loss in Human-AI Interaction. International Journal of Human–Computer Interaction, 1–22. https://doi.org/10.1080/10447318.2024.2405782