Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 07/08/2026 - **Accepted:** 04/09/2026
Vol: 01.04
DOI: 10.66104/4hndmk68
Pages: 1-43

Artificial Intelligence Tools for Productivity Enhancement: A Narrative and Thematic Literature Review

Marcelo Almeida Pereira Guedes
Faculdade Presidente Antônio Carlos, Brazil
Corresponding author: almeidaguedes0@gmail.com

Abstract

Artificial Intelligence (AI), particularly generative AI based on Large Language Models (LLMs), has rapidly transformed the execution of knowledge-intensive activities across multiple domains. This narrative and thematic literature review examines recent evidence on AI-assisted productivity in professional knowledge work, software development, education, scientific research, and organizational settings. The literature was organized thematically to compare productivity metrics, modes of human–AI interaction, task characteristics, user expertise, verification requirements, and governance conditions. The evidence indicates that generative AI can reduce task-completion time or improve output performance in selected activities, but the magnitude and direction of these effects vary across tasks, users, tools, and contexts. Productivity is therefore treated as a multidimensional and conditional outcome rather than as a uniform consequence of AI adoption. Particular attention is given to verification, correction, integration, and governance costs, which may offset initial efficiency gains. The review concludes that AI-assisted productivity is most appropriately understood as a context-dependent result of task–AI fit, human expertise, output reliability, workflow integration, and responsible governance.

Keywords: Artificial Intelligence; Generative AI; Productivity; Large Language Models; Knowledge Work.

1. Introduction

Artificial Intelligence (AI) has emerged as one of the most transformative technologies of the twenty-first century. Although AI research has been developed

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 07/08/2026 - **Accepted:** 04/09/2026
Vol: 01.04
DOI: 10.66104/4hndmk68
Pages: 1-43

for several decades, recent advances in machine learning, deep learning, and Large Language Models (LLMs) have significantly expanded the practical applications of intelligent systems across diverse sectors. Modern AI technologies differ from previous generations because they are capable of generating human-like text, programming code, images, summaries, and analytical content, enabling new forms of interaction between humans and computational systems (Russell & Norvig, 2021; Goodfellow et al., 2016).

The emergence of generative AI represents a major transition in the evolution of artificial intelligence. Unlike traditional AI systems mainly designed for prediction, classification, or automation of predefined tasks, generative models can create new content based on complex patterns learned from large datasets. Foundation models and LLMs have introduced a new paradigm in which AI systems can assist users in writing, programming, research activities, education, and business operations (Bommasani et al., 2021).

The public release of ChatGPT in 2022 accelerated the global adoption of generative AI technologies. In a short period, organizations, universities, software companies, and researchers began incorporating AI assistants into daily workflows. Several competing platforms, including Microsoft Copilot, Google Gemini, Claude, GitHub Copilot, and Notion AI, have expanded the availability of AI-based productivity solutions by providing assistance in document creation, programming, data analysis, communication, and knowledge management.

The increasing adoption of AI tools has generated significant interest regarding their impact on productivity. Recent empirical investigations indicate that AI assistance can reduce task-completion time or improve evaluated output in selected cognitive and knowledge-intensive activities. Experimental evidence has identified performance gains in specific professional writing tasks (Noy & Zhang, 2023), while field and workplace studies have reported improvements under particular task, worker, and organizational conditions (Brynjolfsson et al., 2025; Dell'Acqua et al., 2026). These findings should not be interpreted as evidence of a uniform productivity effect across occupations or contexts. A recent non-peer-reviewed quantitative synthesis likewise reports positive task-level effects while emphasizing substantial heterogeneity across studies (Singh, 2026).

Evidence from AI-assisted software engineering suggests that coding assistants can reduce task-completion time in selected programming activities. However, these effects should not be generalized to all software-development tasks, since performance may vary according to task complexity, developer expertise, project

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 07/08/2026 - **Accepted:** 04/09/2026
Vol: 01.04
DOI: 10.66104/4hndmk68
Pages: 1-43

context, and the amount of verification required for AI-generated code (Peng et al., 2023).

Beyond software engineering, AI tools have also expanded into education, scientific research, and organizational management. In educational environments, generative AI can support explanations, educational materials, feedback, and programming instruction, although the effects vary across settings and outcomes (Rahman et al., 2025; Maier et al., 2026). In scientific research, AI-based systems are being explored for literature analysis, information retrieval, and evidence-synthesis support, with human verification remaining important (de la Torre-López et al., 2023; Clark et al., 2025). Organizations have also adopted AI systems for customer support, information processing, administrative assistance, and knowledge-work activities (Brynjolfsson et al., 2025; Microsoft, 2024).

Despite these advantages, the adoption of AI productivity tools also introduces important challenges. Large Language Models may produce inaccurate information, generate fabricated references, reproduce biases present in training data, and provide inconsistent responses depending on the complexity of the task. Researchers have also raised concerns regarding privacy, copyright, transparency, ethical responsibility, and the possible reduction of human critical thinking when users become excessively dependent on automated systems (Bender et al., 2021; Weidinger et al., 2022).

Another relevant issue is that current research on AI productivity remains fragmented across multiple academic fields. Studies have examined specific applications of AI in software engineering, education, business, and scientific research, but a comprehensive overview integrating evidence from these different domains is still limited. This fragmentation makes it difficult for researchers, organizations, educators, and policymakers to evaluate which AI applications provide meaningful productivity improvements and under what conditions these benefits occur.

A narrative and thematic synthesis of current scientific evidence is useful for integrating findings from these different domains while preserving differences in study design, productivity metrics, and application context. Rather than claiming exhaustive coverage, the present review organizes the literature around recurring themes and contrasts in AI-assisted productivity.

The objective of this study is to conduct a narrative and thematic review of the scientific literature regarding Artificial Intelligence tools designed to enhance

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 07/08/2026 - **Accepted:** 04/09/2026
Vol: 01.04
DOI: 10.66104/4hndmk68
Pages: 1-43

productivity. Specifically, this review aims to:

(i) identify the main AI tools currently applied to productivity improvement;

(ii) analyze their applications in business, education, software development, and scientific research;

(iii) evaluate the benefits and limitations reported in previous studies; and

(iv) discuss future research directions regarding the integration of AI technologies into professional and academic workflows.

By synthesizing recent scientific evidence thematically, this study contributes to a broader understanding of how AI technologies are transforming knowledge work and supporting human productivity under different task, user, and organizational conditions. The findings may assist researchers, educators, software developers, managers, and decision-makers in evaluating both the opportunities and the limitations associated with AI-powered productivity tools.

Although selected studies have reported measurable task- and individual-level productivity gains, historical experience with information technologies suggests caution when extrapolating these effects to sustained organizational or aggregate economic productivity. Brynjolfsson (1993) classically described the 'productivity paradox,' whereby substantial investments in information technology did not immediately translate into economy-wide productivity growth. More recently, Acemoglu (2025) has argued that the macroeconomic effects of AI will depend critically on how the technology is directed—toward automation that displaces labor versus augmentation that complements human capabilities—raising important questions about the long-term distribution of productivity benefits.

2. Literature Review

2.1 Artificial Intelligence Concepts

Artificial Intelligence (AI) refers to the field of computer science dedicated to developing systems capable of performing tasks that traditionally require human intelligence, including reasoning, learning, perception, language understanding, and decision-making. The concept of AI has evolved considerably since its emergence in the mid-twentieth century, moving from rule-based systems toward adaptive

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 07/08/2026 - **Accepted:** 04/09/2026
Vol: 01.04
DOI: 10.66104/4hndmk68
Pages: 1-43

models capable of learning from large volumes of data (Russell & Norvig, 2021).

Early AI approaches were primarily based on symbolic reasoning, where systems operated according to predefined rules created by human experts. Although these methods achieved success in specific domains, they presented limitations when dealing with uncertain environments and complex real-world problems. The development of machine learning introduced a significant change by allowing algorithms to identify patterns from data and improve their performance without requiring explicit programming for every possible situation (Jordan & Mitchell, 2015).

Deep learning further expanded AI capabilities through artificial neural networks composed of multiple computational layers capable of processing complex information. These models achieved significant improvements in areas such as image recognition, speech processing, natural language understanding, and autonomous systems (Goodfellow et al., 2016). The availability of large datasets, increased computational power, and specialized hardware accelerated the adoption of deep learning techniques across academic and industrial applications.

More recently, the emergence of foundation models and Large Language Models (LLMs) has introduced a new stage in AI development. Unlike traditional machine learning systems designed for specific tasks, foundation models can perform multiple activities after being trained on extensive datasets. These systems can generate text, summarize documents, translate languages, write computer code, and assist users in complex knowledge-based activities (Bommasani et al., 2021).

The rapid evolution from traditional AI to generative AI has changed the relationship between humans and computational systems. Current generative AI systems are increasingly used as assistants that can support information access, content generation, and analytical or creative activities, although the extent of augmentation and the need for human oversight vary by task and context (Bommasani et al., 2021; Teubner et al., 2023).

2.2 Evolution of Artificial Intelligence Technologies

The historical development of AI can be divided into several stages, each characterized by different technological approaches and levels of practical adoption. The first period began during the 1950s, when researchers introduced the concept of machines capable of simulating aspects of human intelligence. Early developments focused mainly on symbolic reasoning, problem-solving algorithms, and expert systems designed to reproduce human decision-making processes

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 07/08/2026 - **Accepted:** 04/09/2026
Vol: 01.04
DOI: 10.66104/4hndmk68
Pages: 1-43

(Russell & Norvig, 2021).

During the 1980s and 1990s, expert systems gained attention in business and industrial environments because they could support decision-making in specialized domains. However, these systems required extensive manual knowledge representation and struggled with changing conditions, limiting their scalability. As computational resources improved and digital data became increasingly available, statistical approaches and machine learning methods gradually replaced many rule-based techniques.

The growth of the internet and the expansion of digital information during the 2000s created favorable conditions for data-driven AI development. Machine learning algorithms became increasingly effective because organizations could collect and process large datasets from online platforms, sensors, and information systems. This period also marked the expansion of applications such as recommendation systems, fraud detection, search engines, and predictive analytics.

The emergence of deep learning during the 2010s represented another major transformation. Neural network architectures with multiple layers achieved remarkable improvements in tasks involving images, speech, and natural language processing. These advances were supported by increased computational capacity, graphics processing units (GPUs), and large-scale datasets (Goodfellow et al., 2016).

The introduction of transformer-based architectures and foundation models further accelerated AI development. Large Language Models demonstrated the ability to understand and generate human-like language, enabling applications that extend beyond traditional automation. Tools based on these models can assist users in writing, programming, research, education, and business activities (Bommasani et al., 2021).

The release of ChatGPT in 2022 represented a turning point in public awareness and adoption of generative AI technologies. Generative AI moved from specialized research environments into widespread use, and organizations began exploring these systems as productivity-oriented tools for routine and knowledge-intensive activities (Bick et al., 2026; Microsoft, 2024).

Current AI development increasingly includes multimodal and more general-purpose systems capable of processing or generating multiple forms of information. These developments may broaden the range of tasks supported by AI, but demonstrated productivity effects remain dependent on specific implementations, tasks, and

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 07/08/2026 - **Accepted:** 04/09/2026
Vol: 01.04
DOI: 10.66104/4hndmk68
Pages: 1-43

verification requirements (Bommasani et al., 2021; Ngo et al., 2025).

2.3 Defining and Measuring Productivity in AI-Assisted Work

Productivity is not treated in this review as a single or universally comparable outcome. At its broadest level, productivity refers to the relationship between useful outputs and the resources required to produce them. In AI-assisted work, however, this relationship may be operationalized differently depending on the domain, task, and level of analysis. Relevant outcomes may include reductions in task-completion time, increases in output volume, improvements in output quality or accuracy, reductions in cognitive or procedural effort, and broader organizational effects. Consequently, a reduction in execution time should not automatically be interpreted as a durable productivity gain if it is accompanied by additional verification, correction, integration, or governance costs. Table 1 presents the taxonomy of productivity dimensions and representative metrics adopted in this review.

Table 1. Taxonomy of Productivity Metrics in AI-Assisted Work

Domain	Productivity dimension	Representative metrics	Interpretation in AI-assisted work
Business and organizational work	Efficiency	Task-completion time; processing time; decision speed	Whether AI reduces time for business processes or decisions
Business and organizational work	Output/throughput	Tasks completed; cases processed; output per employee	Whether more useful work is produced per period
Business and organizational work	Quality	Accuracy; customer-service quality; error rates	Whether speed/output is maintained with acceptable or improved quality
Business and organizational work	Organizational performance	Workflow efficiency; coordination; resource utilization	Whether task-level gains translate into broader organizational benefits
Software development	Speed	Development, debugging, or task-completion time	Whether AI reduces time required for defined programming activities

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 07/08/2026 - **Accepted:** 04/09/2026
Vol: 01.04
DOI: 10.66104/4hndmk68
Pages: 1-43

Domain	Productivity dimension	Representative metrics	Interpretation in AI-assisted work
Software development	Output/throughput	Tasks completed; code produced; development output	Whether developers complete more work within a period
Software development	Code quality	Correctness; maintainability; defects; test performance	Whether generated code satisfies technical requirements
Software development	Human effort	Manual coding, debugging, testing, and verification effort	Whether reduced generation effort is offset by review and correction
Education	Performance	Assessment, task, or programming performance	Immediate assisted performance
Education	Efficiency	Task-completion or feedback speed	Time required for educational activities
Education	Learning	Knowledge acquisition; skill development; retention	Durable learning rather than immediate assisted output
Scientific research	Research efficiency	Literature-search, screening, extraction, or synthesis time	Whether AI reduces research-processing effort
Scientific research	Output quality	Reference, extraction, and factual accuracy	Reliability of AI-assisted research outputs
Knowledge work	Speed / output / quality	Writing time; completed tasks; correctness; usefulness	Task-level productivity in text and information work
Knowledge work	Verification and correction	Reviewing, fact-checking, correction, error detection	Resources required after AI generation

This taxonomy highlights an important limitation in comparing productivity findings across domains. A study reporting faster task completion in software development does not measure the same outcome as a study reporting improved learning performance in education or increased output per worker in an

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 07/08/2026 - **Accepted:** 04/09/2026
Vol: 01.04
DOI: 10.66104/4hndmk68
Pages: 1-43

organizational setting. Productivity findings should therefore be interpreted according to the specific metric, task, and level of analysis employed. Throughout this review, evidence is distinguished among speed, output, quality, effort, and broader organizational outcomes rather than being aggregated under a single notion of productivity. Particular attention is also given to verification costs, since time saved during AI-assisted generation may be partially offset by the time required to identify errors, validate information, correct outputs, or integrate AI-generated material into professional workflows.

2.4 Modes of Human–AI Interaction

The productivity effects of Artificial Intelligence depend not only on the capabilities of the technology but also on the role assigned to AI within the work process. To avoid treating all forms of AI use as equivalent, this review distinguishes five modes of human–AI interaction: assistance, augmentation, automation, delegation, and collaboration. These categories represent different distributions of task execution, decision authority, and human oversight. They are used as analytical categories rather than as mutually exclusive technological classes, since the same AI system may operate under different modes depending on how it is integrated into a particular task or workflow. Table 2 summarizes this classification.

Table 2. Modes of Human–AI Interaction in Productivity-Oriented Tasks

Mode	Role of AI	Role of Human	Decision / Control	Typical Example	Productivity Mechanism
Assistance	Provides suggestions, information, explanations, or preliminary outputs	Performs task and decides whether to use AI contribution	Predominantly human	Suggestions, explanations, summaries	Reduces search, preparation, or cognitive effort
Augmentation	Expands the user's existing capability or performance	Remains responsible while using AI to perform more effectively	Human-centered	Writing, analysis, coding, decision support	Improves speed, quality, or capacity

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 07/08/2026 - **Accepted:** 04/09/2026
Vol: 01.04
DOI: 10.66104/4hndmk68
Pages: 1-43

Mode	Role of AI	Role of Human	Decision / Control	Typical Example	Productivity Mechanism
Automation	Executes a defined task or task component	Configures, supervises, or reviews	Primarily system execution with oversight	Classification, code generation, literature screening	Reduces direct execution effort
Delegation	Receives an objective and generates a result on user's behalf	Defines objective and evaluates result	Execution transferred; final responsibility human	Drafting or multi-step task execution	Transfers execution effort but creates verification requirements
Collaboration	Participates iteratively in problem-solving or content production	Iteratively refines prompts, outputs, and decisions	Shared iterative process	Human–AI co-development of text, code, analysis	Combines judgment with rapid generation and feedback

These interaction modes should not be interpreted as a linear progression toward increasing automation. Assistance and augmentation may generate substantial productivity benefits while maintaining strong human control, whereas delegation and automation may reduce direct execution effort but increase the importance of verification and oversight. Collaboration differs from simple assistance because the output emerges through an iterative process in which both human input and AI-generated responses influence subsequent steps. Across these modes, the final productivity effect depends on the balance between the effort reduced by AI and the additional effort required for evaluation, correction, integration, and supervision (Noy & Zhang, 2023; Peng et al., 2023; Dell'Acqua et al., 2026; Ngo et al., 2025).

2.5 Categories of AI Productivity Tools

AI productivity tools can also be classified according to the primary activities they support and the mechanisms through which they may influence productivity. This functional classification is more informative than a simple list of individual applications because tools within the same category frequently share similar tasks, productivity mechanisms, verification requirements, and risks. Table 3 presents the principal categories identified in the reviewed literature.

Table 3. Functional Matrix of AI Productivity Tools

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 07/08/2026 - **Accepted:** 04/09/2026
Vol: 01.04
DOI: 10.66104/4hndmk68
Pages: 1-43

Tool Category	Examples / Technologies	Primary Tasks	Interaction Modes	Productivity Outcomes	Verification Requirement	Evidence
Writing and knowledge assistants	ChatGPT and other LLM-based assistants	Drafting, summarization, rewriting, information processing, ideation	Assistance; augmentation; delegation; collaboration	Completion time; output quality; cognitive effort	Factual accuracy; relevance; originality	Noy & Zhang (2023); Ngo et al. (2025)
Coding assistants	GitHub Copilot and LLM-based coding assistants	Code completion, generation, debugging, explanation, documentation	Assistance; augmentation; partial automation	Development time; task completion; code output	Correctness; security; maintainability; testing	Peng et al. (2023); Ebrahim & Tanner (2025); Maier et al. (2026)
Business and workplace assistants	Generative AI workplace assistants and copilots	Customer support, document processing, communication, information analysis	Assistance; augmentation; partial automation	Throughput; completion time; output quality; worker performance	Accuracy; confidentiality; organizational compliance	Brynjolfsson et al. (2025); Bick et al. (2025); Humlum & Vestergaard (2025)
Research and evidence-synthesis assistants	LLMs and AI-supported review systems	Literature discovery, screening, extraction, summarization, evidence organization	Assistance; augmentation; partial automation	Search/screening time; retrieval efficiency; research effort	Source validity; factual accuracy; reference verification	Cao (2023); de la Torre-López et al. (2023); Gabashvili (2023); Clark et al. (2025)
Educational AI assistants	ChatGPT and AI programming/learning assistants	Explanations, feedback, tutoring, code support	Assistance; augmentation; collaboration	Immediate performance; learning outcomes; completion time	Pedagogical accuracy; student understanding; independent learning	Rahman et al. (2025); Maier et al. (2026)

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 07/08/2026 - **Accepted:** 04/09/2026
Vol: 01.04
DOI: 10.66104/4hndmk68
Pages: 1-43

Tool Category	Examples / Technologies	Primary Tasks	Interaction Modes	Productivity Outcomes	Verification Requirement	Evidence
General-purpose foundation-model systems	Large Language Models and foundation models	Multiple knowledge-work activities	All five modes depending on implementation	Task-dependent	Task-dependent; potentially high for difficult-to-verify outputs	Bommasani et al. (2021); Ngo et al. (2025)

The matrix demonstrates that AI productivity tools differ not only in their technical functions but also in the type of productivity outcome they are expected to influence and in the verification burden imposed on users. Writing and coding assistants, for example, can generate immediately observable outputs, but these outputs must still be evaluated for factual or technical correctness. Research assistants may reduce screening or information-processing effort while imposing particularly important source-verification requirements. Educational assistants present a different challenge because improvements in immediate task performance should not automatically be interpreted as improvements in durable learning. Consequently, comparisons among AI productivity tools should consider the task supported, the interaction mode, the productivity metric, and the verification requirements rather than treating all AI-assisted activities as equivalent (Peng et al., 2023; Cao, 2023; Rahman et al., 2025; Ngo et al., 2025).

2.6 Domain-Specific Risks and Governance Requirements

The risks associated with AI productivity tools are not uniform across application domains. Although hallucination, bias, privacy, security, and overreliance represent recurring concerns, their practical consequences depend on the type of task, the information being processed, and the consequences of an incorrect output. A domain-specific perspective is therefore necessary when evaluating whether an apparent productivity benefit justifies the associated verification and governance requirements. Table 4 relates the principal risks identified in the reviewed literature to their potential consequences and appropriate forms of human or organizational control.

Table 4. Domain-Specific Risks and Governance Requirements for AI-Assisted Productivity

Domain	Principal Risk	Potential Consequence	Productivity Impact	Verification / Governance	Supporting Literature

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 07/08/2026 - **Accepted:** 04/09/2026
Vol: 01.04
DOI: 10.66104/4hndmk68
Pages: 1-43

Domain	Principal Risk	Potential Consequence	Productivity Impact	Verification / Governance	Supporting Literature
Software development	Incorrect or insecure generated code	Defects, vulnerabilities, inappropriate implementation	Generation savings may be offset by debugging/testing	Code review; testing; security analysis; developer validation	Peng et al. (2023); Ebrahim & Tanner (2025)
Scientific research	Fabricated, inaccurate, or unverifiable information/references	Invalid citations; inaccurate evidence synthesis	Faster retrieval/drafting may create verification costs	Check original sources; researcher oversight; extraction validation	Cao (2023); de la Torre-López et al. (2023); Clark et al. (2025)
Education	Overreliance and inaccurate explanations	Superficial completion or weaker independent learning	Immediate performance may not equal durable learning	Pedagogical supervision; assessment design; verify learning	Rahman et al. (2025); Maier et al. (2026)
Business and organizational work	Confidentiality, privacy, inaccurate information	Data exposure; incorrect decisions; reputational risk	Efficiency may be offset by governance/compliance costs	Data governance; human oversight; access controls; training	Ngo et al. (2025); Microsoft (2024); UNESCO (2021)
Knowledge work and content creation	Hallucination, bias, inaccurate content	Misleading documents; low-quality outputs	Rapid production may increase review burden	Fact-checking; editorial review; responsible-use procedures	Bender et al. (2021); Weidinger et al. (2022); Ngo et al. (2025)
AI-assisted decision support	Inaccurate or biased recommendations; automation reliance	Poor decisions; reduced scrutiny	Faster analysis may not improve decision quality	Human accountability; critical-output validation; transparency	Floridi & Cowls (2019); UNESCO (2021); Weidinger et al. (2022)

The domain-specific comparison shows why risk should be incorporated into

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 07/08/2026 - **Accepted:** 04/09/2026
Vol: 01.04
DOI: 10.66104/4hndmk68
Pages: 1-43

productivity analysis rather than treated as a separate ethical issue. An AI-generated error does not impose the same cost in every context. An inaccurate suggestion in an early writing draft may be relatively inexpensive to correct, whereas an insecure code recommendation, an invalid scientific reference, or an inappropriate organizational decision may require substantial corrective effort or generate consequences beyond the immediate task. Verification cost therefore varies according to both the probability of error and the consequences of accepting an incorrect output (Cao, 2023; Peng et al., 2023; Weidinger et al., 2022).

This relationship also reinforces the conceptual framework proposed in this review. Productivity gains are more likely to translate into net benefits when AI outputs can be evaluated efficiently and the consequences of errors are manageable. In high-stakes or difficult-to-verify domains, stronger human oversight and governance may be required, increasing the resources necessary to obtain a reliable final output. Responsible AI governance should therefore be interpreted not only as an ethical requirement but also as a component of the total cost of AI-assisted productivity (Floridi & Cowls, 2019; UNESCO, 2021; Ngo et al., 2025).

2.7 AI Tools for Productivity Enhancement

The rapid advancement of generative AI has resulted in numerous productivity-oriented applications that support users through content generation, recommendations, information processing, and partial automation. The contribution of these systems to productivity varies according to the task, interaction mode, output reliability, and the amount of human evaluation required (Ngo et al., 2025; Teubner et al., 2023).

2.7.1 AI Writing and Knowledge Assistants

Text-based AI assistants represent a widely used category of generative AI productivity tools. Applications such as ChatGPT and related LLM-based systems can generate text, summarize documents, answer questions, translate information, and support knowledge-work activities (Bick et al., 2026; Ngo et al., 2025).

These systems have been increasingly used in professional communication, academic writing, customer support, and knowledge management. Experimental studies indicate that generative AI can improve productivity in writing-related tasks by reducing completion time and helping users overcome difficulties related to information organization and content generation (Noy & Zhang, 2023).

However, the effectiveness of AI writing assistants depends strongly on human

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 07/08/2026 - **Accepted:** 04/09/2026
Vol: 01.04
DOI: 10.66104/4hndmk68
Pages: 1-43

supervision. Although these tools can accelerate drafting processes, they may generate incorrect information, unsupported claims, or misleading responses. Therefore, users must evaluate AI-generated content before applying it in professional or academic contexts (Ngo et al., 2025).

2.7.2 AI Tools for Software Development

Software engineering has become an important domain for investigating AI-assisted productivity. Coding assistants such as GitHub Copilot and similar generative AI tools can provide code completion, generate programming functions, explain existing code, suggest improvements, assist debugging, and create documentation. Evidence from specific programming tasks indicates that AI assistance can reduce completion time and support developer efficiency, particularly in relatively well-defined coding activities. However, the available findings should not be interpreted as demonstrating equivalent productivity gains across all programming tasks or development environments (Peng et al., 2023).

Nevertheless, AI-assisted programming introduces new challenges. Generated code may contain incorrect, insecure, or inappropriate solutions, so developers must retain technical judgment and review, test, and validate AI-generated suggestions before integration into production systems (Peng et al., 2023; Ebrahim & Tanner, 2025).

2.7.3 AI Tools for Business and Office Productivity

Organizations have increasingly incorporated AI assistants into administrative and business processes. Tools such as Microsoft Copilot, Google Workspace AI features, and Notion AI aim to improve workplace productivity by assisting with document creation, meeting summaries, data analysis, scheduling, and information organization. Industry reports indicate that the adoption of these tools is advancing rapidly across knowledge-intensive occupations, with a growing share of employees already using generative AI in their daily work (Microsoft, 2024).

Research conducted in workplace settings indicates that generative AI can improve efficiency in selected knowledge-intensive tasks, including by reducing the time required for some routine or information-processing activities. The magnitude of these benefits varies across workers and tasks, and task-level efficiency gains do not necessarily translate directly into sustained organizational productivity. AI systems may also support decision-making by processing and organizing information, although the usefulness of such support depends on output reliability,

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 07/08/2026 - **Accepted:** 04/09/2026
Vol: 01.04
DOI: 10.66104/4hndmk68
Pages: 1-43

human evaluation, and appropriate workflow integration (Brynjolfsson et al., 2025).

However, organizational adoption requires appropriate policies concerning data protection, employee training, access, and responsible use. Without adequate governance, AI implementation may create confidentiality, reliability, and unequal-adoption risks that can offset or complicate expected productivity benefits (Humlum & Vestergaard, 2025; Ngo et al., 2025; UNESCO, 2021).

2.7.4 AI Tools for Education and Scientific Research

Education and scientific research have also experienced growing use of generative AI tools. In these domains, AI can support explanations, summarization, educational content, programming assistance, literature discovery, and evidence processing, but the value of such support depends on pedagogical context, output reliability, and human verification (Rahman et al., 2025; de la Torre-López et al., 2023; Clark et al., 2025).

In programming education, systematic reviews indicate that AI assistants can support learning and task performance through mechanisms such as personalized guidance, code generation, explanations, and immediate feedback (Rahman et al., 2025). However, reported effects vary across educational settings, tools, outcome measures, and study designs. Evidence concerning faster task performance should also be distinguished from evidence of durable learning or skill acquisition, since improvements in immediate performance do not necessarily imply equivalent long-term educational gains (Maier et al., 2026).

In scientific research, AI systems have been explored for literature discovery, article summarization, and evidence synthesis. These applications can reduce the time required for information retrieval and analysis, although researchers must carefully validate AI-generated information because language models may produce inaccurate references or unsupported conclusions (Gabashvili, 2023).

2.8 Benefits of AI Productivity Tools

The literature identifies several potential benefits associated with AI adoption in productivity-oriented activities, but the magnitude and durability of these benefits vary across tasks, users, and study designs (Noy & Zhang, 2023; Brynjolfsson et al., 2025; Ngo et al., 2025).

First, AI tools can reduce the time required to complete certain repetitive and information-intensive tasks. Evidence of time savings has been reported in selected activities involving professional writing, customer support, programming, and

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 07/08/2026 - **Accepted:** 04/09/2026
Vol: 01.04
DOI: 10.66104/4hndmk68
Pages: 1-43

information processing. The size of these gains varies according to the task, tool, user, and evaluation setting. Moreover, faster initial task execution should not automatically be interpreted as a net productivity improvement when substantial time is subsequently required to verify, correct, test, or integrate AI-generated outputs (Noy & Zhang, 2023; Brynjolfsson et al., 2025; Peng et al., 2023).

Second, AI assistants can improve access to task-relevant information by providing explanations, summaries, and recommendations in natural language. This form of support may reduce search or preparation effort for some users, although usefulness remains dependent on output accuracy and the user's ability to evaluate the information provided (Ngo et al., 2025; Teubner et al., 2023).

Third, generative AI can support ideation and problem-solving by producing alternative drafts, suggestions, or candidate solutions. These capabilities are more appropriately interpreted as forms of assistance or augmentation than as evidence that AI replaces human creativity or judgment (Teubner et al., 2023; Ngo et al., 2025).

Finally, some empirical studies suggest that less experienced users may obtain comparatively larger benefits from AI assistance in particular task settings, as AI systems can provide guidance and make task-relevant knowledge more readily accessible (Noy & Zhang, 2023; Brynjolfsson et al., 2025). Field experimental evidence also indicates heterogeneous effects according to worker expertise and task characteristics (Dell'Acqua et al., 2026). These findings should not be generalized to all occupations or levels of expertise, since experienced users may retain advantages in evaluating outputs, identifying errors, managing complex tasks, and integrating AI-generated material into professional workflows.

2.9 Limitations and Ethical Challenges

Despite their potential benefits, AI productivity tools present significant limitations. One recurring concern is hallucination or the generation of plausible but incorrect information, which can impose substantial verification costs in academic, professional, and decision-making contexts (Cao, 2023; Ngo et al., 2025).

Another challenge involves algorithmic bias. Since AI models learn from large datasets produced by humans, they may reproduce existing social and informational biases. These limitations can affect the fairness and reliability of AI-generated outputs (Bender et al., 2021).

Privacy and data protection represent additional concerns because AI productivity

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 07/08/2026 - **Accepted:** 04/09/2026
Vol: 01.04
DOI: 10.66104/4hndmk68
Pages: 1-43

systems may process information that is confidential, proprietary, or otherwise sensitive. These risks make organizational policies, access controls, and responsible-use practices important components of AI governance (Ngo et al., 2025; UNESCO, 2021; Weidinger et al., 2022).

Ethical considerations also include transparency, accountability, and responsible use. Researchers argue that AI systems should be developed and deployed according to principles that prioritize human oversight, safety, fairness, and social benefit (Floridi & Cowls, 2019; UNESCO, 2021).

Therefore, the reviewed literature does not support treating AI productivity tools as complete replacements for human expertise. Their contribution is more appropriately understood through assistance, augmentation, selective automation, delegation, or collaboration while maintaining human evaluation and responsibility, particularly where outputs are difficult to verify or errors carry significant consequences (Peng et al., 2023; Dell'Acqua et al., 2026; Ngo et al., 2025).

3. Methodology / Materials and Methods

This study was conducted as a narrative and thematic literature review to identify, organize, and interpret recent scientific evidence concerning the use of Artificial Intelligence tools for productivity enhancement. The review was designed to compare recurring patterns, contrasts, and limitations across heterogeneous research domains rather than to claim exhaustive retrieval or to estimate a pooled effect size. This framing is appropriate because the available literature includes controlled experiments, field studies, surveys, systematic reviews, meta-analyses, conference papers, preprints, and theoretical or institutional sources that address different productivity outcomes and levels of analysis.

3.1 Literature Search Strategy

The literature search used Google Scholar, IEEE Xplore, ACM Digital Library, ScienceDirect, SpringerLink, and Scopus. The literature considered primarily covered the period from 2020 to 2026. Search terms included "Artificial Intelligence," "Generative AI," "AI productivity tools," "Large Language Models," "ChatGPT," "GitHub Copilot," "AI-assisted productivity," "workflow automation," and "knowledge work." Terms were combined as appropriate to identify literature on AI-supported productivity in business, education, software development, scientific research, and knowledge work. Because the review is narrative and thematic rather than a

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 07/08/2026 - **Accepted:** 04/09/2026
Vol: 01.04
DOI: 10.66104/4hndmk68
Pages: 1-43

PRISMA-style systematic review, the search was used to construct a focused multidisciplinary evidence base rather than to claim exhaustive retrieval of all records.

3.2 Inclusion, Exclusion, and Source Classification

Priority was given to English-language publications from 2020 to 2026 addressing AI tools, generative AI, Large Language Models, or AI-assisted productivity and reporting applications, outcomes, limitations, or adoption effects relevant to productivity. Peer-reviewed journal articles and scientific conference papers were prioritized. Duplicate publications, non-scientific blogs, news reports, sources unrelated to productivity, and publications focused exclusively on technical AI development without practical application were excluded from the analytical synthesis. Selected preprints and working papers were retained when they addressed recent developments for which peer-reviewed evidence was limited, but their publication status was considered during evidence appraisal.

The final reference list contains 29 publications. Of these, 15 publications were classified as the analytical corpus because they contributed directly to the thematic synthesis of productivity outcomes, adoption, reliability, learning, software development, or evidence-synthesis applications. The remaining 14 references were used as contextual, theoretical, ethical, historical, or institutional sources and were not treated as equivalent analytical studies. This distinction prevents the total bibliography from being presented as if every citation represented an included empirical productivity study.

3.3 Study Selection and Thematic Analysis

The literature screening, eligibility assessment, study classification, data extraction, quality appraisal, and thematic coding were conducted by a single reviewer. Consequently, no inter-rater reliability statistic was calculated. The analytical categories were applied consistently across the selected publications, but the absence of independent duplicate screening and coding may have increased susceptibility to subjective judgments in study selection, classification, and interpretation. This methodological characteristic is one reason the present study is framed as a narrative and thematic literature review rather than as a systematic review.

The analytical process involved four stages: (1) identifying publications relevant to AI-assisted productivity; (2) distinguishing analytical studies from contextual and theoretical references; (3) extracting information on study design, domain, AI

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 07/08/2026 - **Accepted:** 04/09/2026
Vol: 01.04
DOI: 10.66104/4hndmk68
Pages: 1-43

technology, productivity outcome, main findings, publication status, and limitations; and (4) organizing the evidence into recurring themes and contrasts. The thematic synthesis focused particularly on task structure, user expertise, modes of human–AI interaction, measured versus perceived productivity, verification requirements, organizational conditions, and short-term versus long-term effects.

3.4 Characteristics of the Analytical Literature

To improve the transparency of the thematic synthesis, Table 5 summarizes the main characteristics of the publications that contributed directly to the analytical discussion. Because the literature encompasses experimental studies, observational research, systematic reviews, meta-analyses, and surveys, the reported samples refer either to participants or to the number of studies included in evidence syntheses, as applicable. The table also distinguishes the productivity-related outcomes examined in each publication and the principal limitations relevant to the interpretation of the evidence.

Table 5. Characteristics of the Analytical Literature

Study	Domain	Sample / Evidence Base	Study Type	AI Tool	Task / Focus	Productivity Metric	Main Finding	Main Limitation
Noy & Zhang (2023)	Professional writing	453 professionals	Preregistered experiment	Generative AI	Writing tasks	Time; output quality	Faster completion and improved evaluated output	Short, defined writing tasks
Brynjolfsson et al. (2025)	Customer support	5,179 agents	Field study	Generative AI assistant	Customer support	Issues/hour; quality	Average gains with larger benefits for less experienced workers	Single organizational setting

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 07/08/2026 - **Accepted:** 04/09/2026
Vol: 01.04
DOI: 10.66104/4hndmk68
Pages: 1-43

Study	Domain	Sample / Evidence Base	Study Type	AI Tool	Task / Focus	Productivity Metric	Main Finding	Main Limitation
Peng et al. (2023)	Software development	Recruited developers	Controlled experiment	GitHub Copilot	Programming task	Completion time	AI-assisted group completed task faster	Single defined task; preprint
Dell'Acqua et al. (2026)	Consulting knowledge work	758 consultants	Preregistered field experiment	GPT-4-based AI	Professional tasks	Completion; speed; quality	Benefits within capability frontier; worse performance outside frontier	Task-specific frontier effects
Bick et al. (2026)	Workforce adoption	Large worker survey	Observational survey	Generative AI	Adoption/use	Use and diffusion	Rapid adoption across workers	Adoption is not direct productivity
Humlum & Vestergaard (2025)	Occupational adoption	Large Danish worker survey	Survey / working paper	ChatGPT	Adoption patterns	Use/adoption	Substantial adoption differences across workers/occupations	Indirect productivity evidence; working paper
Rahman et al. (2025)	Programming education	32 studies in Results	Systematic review/meta-analysis	AI tools	Programming learning	Performance; completion; learning	Positive performance evidence with heterogeneous outcomes	Abstract/results count discrepancy; heterogeneity
Maier et al. (2026)	Programming	23 studies; 27 effect sizes	Meta-analysis / preprint	Generative AI coding tools	Programming productivity/learning	Productivity; learning	Modest productivity effect; learning effect uncertain	Preprint; heterogeneity

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 07/08/2026 - **Accepted:** 04/09/2026
Vol: 01.04
DOI: 10.66104/4hndmk68
Pages: 1-43

Study	Domain	Sample / Evidence Base	Study Type	AI Tool	Task / Focus	Productivity Metric	Main Finding	Main Limitation
Ebrahim & Tanner (2025)	Software engineering	26 articles	Systematic literature review	ChatGPT	Developer support	Synthesized productivity effects	Benefits with reliability/security concerns	Heterogeneous primary studies
Ngo et al. (2025)	Cross-domain ChatGPT use	33 empirical studies	Systematic review	ChatGPT	Limitations /opportunities	Indirect productivity implications	Identifies limitations and opportunities	Broad applications/outcomes
Cao (2023)	Scientific research	Multiple disciplines	Empirical evaluation / preprint	GPT-4	Scientific references	Reference validity	Substantial reliability limitations	Indirect productivity measure; preprint
de la Torre-López et al. (2023)	Evidence synthesis	34 primary studies	Review/survey	AI for SLR automation	Review automation	Screening/automation potential	AI can support stages of literature review	Heterogeneous techniques/studies
Gabashvili (2023)	Research applications	11 reviews	Review of reviews / preprint	ChatGPT	Research applications	Synthesized benefits/limitations	Broad potential with early evidence	Secondary literature; preprint
Clark et al. (2025)	Evidence synthesis	19 studies	Systematic review	Generative AI	Evidence synthesis	Accuracy; errors; efficiency	High error burden requires human oversight	Underlying studies heterogeneous
Singh (2026)	Worker productivity	Multiple studies	Systematic review/quantitative synthesis / SSRN	Generative AI	Worker productivity	Productivity outcomes	Contrasts task-level gains with broader evidence	Working paper; heterogeneous evidence

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 07/08/2026 - **Accepted:** 04/09/2026
Vol: 01.04
DOI: 10.66104/4hndmk68
Pages: 1-43

3.5 Quality Appraisal of the Evidence

Given the methodological heterogeneity of the literature, the publications included in the analytical synthesis were not assumed to provide equivalent levels of evidence. A structured qualitative appraisal was therefore conducted considering study design, sample or evidence base, presence of a comparison or control condition when applicable, directness of the productivity measure, publication status, generalizability, and major sources of potential bias. Because the review encompasses experimental studies, observational research, surveys, systematic reviews, meta-analyses, and selected preprints, the appraisal was intended to support evidence weighting rather than to generate a single numerical quality score. The resulting assessment is presented in Table 6.

Table 6. Qualitative Quality Appraisal of the Evidence Included in the Thematic Synthesis

Study	Design	Comparison / Control	Directness	Publication Status	Main Limitation	Evidence Strength
Noy & Zhang (2023)	Preregistered experiment	Yes	Direct time/quality	Peer-reviewed	Short professional writing tasks	High
Brynjolfsson et al. (2025)	Field study	Comparison/staggered access	Direct workplace productivity	Peer-reviewed	Single organization; expertise effects	High
Dell'Acqua et al. (2026)	Preregistered field experiment	Yes	Direct speed/quality	Peer-reviewed	Capability-frontier dependence	High
Peng et al. (2023)	Controlled experiment	Yes	Direct completion time	Preprint	Single programming task	Moderate
Bick et al. (2026)	Large observational survey	No experimental control	Indirect adoption/use	Peer-reviewed version available	Adoption ≠ productivity	Moderate
Humlum & Vestergaard (2025)	Large survey	No experimental control	Indirect adoption/use	Working paper	Not direct productivity	Moderate

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 07/08/2026 - **Accepted:** 04/09/2026
Vol: 01.04
DOI: 10.66104/4hndmk68
Pages: 1-43

Study	Design	Comparison / Control	Directness	Publication Status	Main Limitation	Evidence Strength
Rahman et al. (2025)	Systematic review/meta-analysis	Across studies	Direct/indirect educational outcomes	Peer-reviewed	Heterogeneity; reporting discrepancy	High
Maier et al. (2026)	Meta-analysis	Across studies	Direct programming productivity/learning	Preprint	Heterogeneity	Moderate
Ebrahim & Tanner (2025)	Systematic literature review	N/A	Synthesized/indirect	Conference	Heterogeneous primary studies	Moderate
Ngo et al. (2025)	Systematic review	N/A	Synthesized/indirect	Peer-reviewed	Broad applications/outcomes	Moderate–High
Cao (2023)	Empirical evaluation	No conventional control	Indirect reliability/verification burden	Preprint	Productivity not measured directly	Low–Moderate
de la Torre-López et al. (2023)	Review/survey	N/A	Indirect automation potential	Peer-reviewed	Heterogeneous evidence	Moderate
Gabashvili (2023)	Review of reviews	N/A	Indirect	Preprint	Early/secondary literature	Low–Moderate
Clark et al. (2025)	Systematic review	N/A	Accuracy/errors/efficiency	Peer-reviewed	Underlying-study bias/heterogeneity	Moderate–High
Singh (2026)	Systematic/quantitative synthesis	Across studies	Direct/indirect	SSRN working paper	Heterogeneous; not peer-reviewed	Moderate

Evidence strength was interpreted qualitatively rather than as a formal numerical risk-of-bias score. High-strength evidence was assigned primarily to peer-reviewed experimental, field-experimental, or rigorous evidence-synthesis research with relatively direct productivity measures. Moderate-strength evidence generally included observational studies, surveys, conference publications, or evidence

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 07/08/2026 - **Accepted:** 04/09/2026
Vol: 01.04
DOI: 10.66104/4hndmk68
Pages: 1-43

syntheses with substantial heterogeneity. Low-to-moderate assessments were applied where productivity was measured only indirectly, the evidence was based predominantly on secondary literature, or the publication had not undergone peer review. These categories were used to guide interpretation and should not be understood as substitutes for design-specific validated risk-of-bias instruments.

4. Results

The narrative and thematic synthesis indicates that Artificial Intelligence tools can support productivity in several knowledge-intensive domains, although the evidence does not support a uniform effect across tasks, users, tools, or organizational contexts. Across the analyzed literature, AI is more consistently described as a technology for assisting or augmenting human activity through information processing, content generation, task support, and partial automation rather than as a complete substitute for human expertise. The productivity outcomes associated with these mechanisms depend on task characteristics, user expertise, output reliability, verification requirements, and the conditions under which AI is integrated into existing workflows.

Empirical evidence is comparatively stronger for selected writing, customer-service, programming, and other information-intensive tasks in which outcomes such as completion time or output quality can be measured directly (Noy & Zhang, 2023; Brynjolfsson et al., 2025; Dell'Acqua et al., 2026). Even within these domains, however, reported gains vary across experimental conditions and should not be extrapolated automatically to long-term individual, organizational, or economy-wide productivity. The interpretation of these findings must therefore consider both the immediate benefits of AI assistance and the additional costs associated with verification, correction, integration, training, and governance.

4.1 Structured versus Open-Ended Tasks

A central pattern emerging from the reviewed literature is that the productivity effects of generative AI vary according to task characteristics. Evidence of measurable gains is comparatively clearer in activities with identifiable objectives, relatively short execution cycles, and outputs that can be evaluated against recognizable criteria. Professional writing, customer-support activities, and selected programming tasks illustrate contexts in which AI assistance has been associated with reductions in task-completion time, increases in throughput, or improvements in evaluated output quality (Noy & Zhang, 2023; Brynjolfsson et al., 2025; Peng et al.,

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 07/08/2026 - **Accepted:** 04/09/2026
Vol: 01.04
DOI: 10.66104/4hndmk68
Pages: 1-43

2023).

The importance of task characteristics is also evident in experimental research showing that the effectiveness of generative AI can change substantially depending on whether the task falls within or outside the capabilities of the AI system. Dell'Acqua et al. (2026) found that AI assistance could improve performance on tasks within the technological capability frontier while producing less favorable outcomes when users relied on AI in tasks beyond that frontier. This finding reinforces the need to avoid interpreting AI productivity as a uniform effect independent of task requirements (Dell'Acqua et al., 2026).

More structured tasks may also facilitate verification because users have clearer criteria for determining whether an output is acceptable. In software development, for example, generated code can frequently be executed, tested, or inspected against technical requirements, although additional evaluation remains necessary because AI-generated code may contain errors or inappropriate solutions. Evidence concerning coding assistants therefore suggests that reduced completion time must be considered together with the human effort required to inspect and validate generated code (Peng et al., 2023; Ebrahim & Tanner, 2025).

The situation becomes more complex in open-ended or evidence-sensitive activities. In scientific research, generative AI can assist literature discovery, screening, organization, and evidence synthesis, but the usefulness of these applications depends heavily on the accuracy and reliability of the generated information. Research evaluating AI-generated scientific references has demonstrated that apparently plausible outputs may include inaccurate or unverifiable information, making human validation essential (Cao, 2023; de la Torre-López et al., 2023).

Taken together, these findings suggest that productivity should not be inferred solely from the amount of material generated or from the speed of initial task completion. In this review, the relevant outcome is whether AI reduces the overall effort required to obtain an acceptable result after generation, evaluation, correction, and integration are considered. The balance between initial assistance and subsequent verification is particularly important in tasks where errors are difficult to identify or may produce significant professional consequences (Cao, 2023; Ngo et al., 2025; Dell'Acqua et al., 2026).

4.2 Novice versus Expert Users

A second important contrast concerns user expertise. Empirical studies indicate that

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 07/08/2026 - **Accepted:** 04/09/2026
Vol: 01.04
DOI: 10.66104/4hndmk68
Pages: 1-43

the benefits of generative AI are not distributed uniformly across workers. In a workplace study involving customer-support agents, Brynjolfsson et al. (2025) reported larger productivity improvements among less experienced and lower-skilled workers than among more experienced workers. Experimental evidence from professional writing tasks similarly suggests that users with initially weaker performance may obtain comparatively larger benefits from generative AI assistance (Noy & Zhang, 2023; Brynjolfsson et al., 2025).

One possible interpretation of this pattern is that generative AI can provide less experienced users with immediate access to examples, recommendations, explanations, and task-relevant information that would otherwise depend more strongly on accumulated experience. In this sense, AI assistance may partially reduce initial performance differences in particular tasks by making useful knowledge or procedural guidance more readily available (Brynjolfsson et al., 2025).

However, comparatively larger gains among less experienced users do not imply that expertise becomes unnecessary. AI-generated outputs must still be evaluated for correctness, relevance, security, and suitability for the specific context. This is particularly evident in software engineering, where coding assistants can accelerate some programming activities while developers remain responsible for reviewing and validating generated solutions before integration into production environments (Peng et al., 2023; Ebrahim & Tanner, 2025).

Dell'Acqua et al. (2026) further demonstrate why expertise remains important when AI capabilities do not align with task requirements. When a task falls outside the effective capability frontier of the system, users must be able to recognize limitations in the AI-generated response rather than relying on the system automatically. Expertise can therefore contribute not only to performing the task itself but also to deciding when AI assistance should be accepted, questioned, corrected, or rejected (Dell'Acqua et al., 2026).

The comparison between novice and expert users therefore suggests a conditional relationship between expertise and AI-assisted productivity. Less experienced users may benefit more from guidance and reduced entry barriers in some tasks, whereas experienced users may retain advantages in evaluation, error detection, contextual judgment, and management of complex situations. The productivity effect consequently depends on the interaction between AI capability, task requirements, and the user's ability to evaluate the generated output (Brynjolfsson et al., 2025; Dell'Acqua et al., 2026).

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 07/08/2026 - **Accepted:** 04/09/2026
Vol: 01.04
DOI: 10.66104/4hndmk68
Pages: 1-43

4.3 Automation versus Augmentation

The reviewed literature also indicates that productivity effects should be interpreted according to the role performed by AI within the human–technology relationship. Automation refers to situations in which an AI system performs a task or task component that would otherwise require direct human execution, whereas augmentation occurs when AI assists human performance while the user retains responsibility for evaluation, selection, interpretation, or decision-making. Current generative AI applications frequently combine these two mechanisms rather than operating exclusively as either complete automation or simple assistance (Noy & Zhang, 2023; Peng et al., 2023; de la Torre-López et al., 2023).

Examples of partial automation include generating an initial text draft, producing code, summarizing documents, organizing information, or supporting stages of literature screening. Nevertheless, the reviewed studies frequently preserve an important human role after the initial output has been generated. In professional writing, programming, and scientific research, users remain responsible for determining whether generated material is accurate, relevant, appropriate, and suitable for the intended purpose (Noy & Zhang, 2023; Peng et al., 2023; Cao, 2023).

This distinction is particularly relevant because automation of one stage of a workflow does not necessarily eliminate human effort. Instead, effort may be redistributed from initial production toward supervision, validation, correction, testing, or integration. Research concerning AI-assisted programming and scientific evidence synthesis illustrates this shift: AI may accelerate generation or screening while simultaneously creating a requirement for careful human verification of the resulting output (Peng et al., 2023; de la Torre-López et al., 2023; Ngo et al., 2025).

For this reason, the present review distinguishes automation from augmentation when interpreting productivity effects. A reduction in direct execution effort represents only one component of the productivity outcome. If a substantial amount of time is subsequently required to identify errors, verify information, correct generated material, or integrate it into an existing workflow, part of the initial efficiency gain may be offset (Cao, 2023; Ngo et al., 2025).

AI assistance benefit → reduced execution effort → verification and integration requirements → net productivity outcome.

Under this framework, AI-assisted productivity depends on the total combination of human and technological effort required to obtain an acceptable output rather than

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 07/08/2026 - **Accepted:** 04/09/2026
Vol: 01.04
DOI: 10.66104/4hndmk68
Pages: 1-43

simply on the amount of work initially performed by the AI system. The reviewed evidence therefore more strongly supports a model of human–AI augmentation and selective automation than a general model of complete replacement of human expertise (Peng et al., 2023; Ebrahim & Tanner, 2025; Ngo et al., 2025).

4.4 Measured versus Perceived Productivity

Another important distinction concerns how productivity is measured. The literature considered in this review includes direct performance measures as well as evidence concerning adoption, perceived usefulness, and potential productivity benefits. These categories provide different forms of evidence and should not be treated as equivalent.

Experimental and field-based studies provide comparatively direct productivity evidence when they measure outcomes such as task-completion time, output quality, or work performed within a defined period. Noy and Zhang (2023), for example, evaluated completion time and quality in professional writing tasks. Brynjolfsson et al. (2025) examined productivity outcomes in a real customer-support environment, while Peng et al. (2023) evaluated completion time in a defined software-development task. These studies therefore provide evidence concerning observable task-level performance rather than only users' perceptions of AI usefulness (Noy & Zhang, 2023; Brynjolfsson et al., 2025; Peng et al., 2023).

By contrast, research concerning the adoption of generative AI addresses different questions. Bick et al. (2026) examined the rapid diffusion of generative AI among workers, while Humlum and Vestergaard (2025) investigated differences in ChatGPT adoption across occupations and worker characteristics. Such studies are valuable for understanding diffusion, access, professional differences, and organizational conditions, but the use or adoption of an AI tool does not by itself demonstrate that the tool has increased productivity (Bick et al., 2026; Humlum & Vestergaard, 2025).

This distinction is important because perceived efficiency may not correspond exactly to realized productivity. An AI-generated output may appear to save time during initial production but subsequently require verification or correction. Conversely, a technology may initially impose learning and integration costs that are not reflected in short-duration evaluations. Research documenting reliability limitations and inaccurate outputs therefore provides an important counterpoint to studies based primarily on speed or adoption (Cao, 2023; Ngo et al., 2025).

Consequently, productivity claims in this review are interpreted according to the type

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 07/08/2026 - **Accepted:** 04/09/2026
Vol: 01.04
DOI: 10.66104/4hndmk68
Pages: 1-43

of outcome reported. Measured reductions in completion time, increases in throughput, or improvements in independently evaluated output provide stronger evidence of task-level productivity effects than adoption rates or self-reported usefulness alone. Adoption evidence remains relevant, but it is treated as evidence of diffusion and use rather than as direct proof of productivity improvement (Noy & Zhang, 2023; Brynjolfsson et al., 2025; Bick et al., 2026; Humlum & Vestergaard, 2025).

This distinction also limits the extent to which findings can be generalized from individuals to organizations. Improvements measured for a worker completing a particular task do not automatically establish that an organization as a whole has become more productive. Organizational outcomes additionally depend on workflow integration, training, coordination, governance, security, and the distribution of benefits and costs across multiple activities (Ngo et al., 2025; Microsoft, 2024).

4.5 Short-Term Gains versus Long-Term Productivity

The temporal dimension introduces another important distinction in the interpretation of AI productivity research. Much of the available empirical evidence concerns immediate task performance, including whether users complete an activity faster, produce higher-quality outputs, or perform more work during a defined experimental or workplace period. Such findings provide evidence of short-term or task-level effects but do not automatically establish sustained productivity improvements over longer periods (Noy & Zhang, 2023; Brynjolfsson et al., 2025; Dell'Acqua et al., 2026).

This distinction is especially relevant in education. AI assistance may improve immediate performance by providing explanations, examples, code suggestions, or personalized guidance, but assisted task performance should be distinguished from durable learning and independent skill development. Systematic reviews and meta-analytic evidence indicate that educational outcomes vary across studies, tools, tasks, and evaluation methods, making it inappropriate to treat immediate task improvement as equivalent to long-term learning (Rahman et al., 2025; Maier et al., 2026).

Similar considerations apply to professional environments. A worker may complete a task more quickly with AI assistance, but organizations may simultaneously incur costs associated with employee training, verification procedures, workflow redesign, data protection, governance, and integration with existing systems. Therefore, short-term performance improvements observed at the individual or task level cannot

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 07/08/2026 - **Accepted:** 04/09/2026
Vol: 01.04
DOI: 10.66104/4hndmk68
Pages: 1-43

automatically be extrapolated to sustained organization-wide productivity (Ngo et al., 2025; Microsoft, 2024).

The distinction between immediate technological capability and broader productivity outcomes is consistent with the historical productivity paradox described by Brynjolfsson (1993). Earlier investments in information technology did not always produce immediate economy-wide productivity improvements because technological benefits could depend on complementary organizational investments and changes in working practices. Although generative AI differs substantially from earlier information technologies, the productivity paradox provides a useful theoretical caution against assuming that rapid technological adoption will automatically generate equivalent aggregate productivity growth (Brynjolfsson, 1993).

More recent economic analysis provides a similar reason for distinguishing task-level gains from broader economic consequences. Acemoglu (2025) argues that the macroeconomic impact of AI depends on the range and economic importance of tasks affected by the technology and on whether AI development primarily contributes to useful automation or to forms of augmentation that complement human work. Consequently, evidence that AI improves performance in particular tasks cannot by itself determine the magnitude of long-term economy-wide productivity effects (Acemoglu, 2025).

Taken together, the reviewed evidence supports a multilevel interpretation of productivity. At the task level, generative AI can produce measurable gains under specific conditions (Noy & Zhang, 2023; Peng et al., 2023; Dell'Acqua et al., 2026). At the individual level, outcomes vary according to user expertise and the ability to evaluate AI-generated outputs (Brynjolfsson et al., 2025; Dell'Acqua et al., 2026). At the organizational level, outcomes additionally depend on integration, governance, training, and complementary changes in work processes (Ngo et al., 2025; Microsoft, 2024). At the broader economic level, the translation of these effects into sustained productivity growth remains more uncertain (Brynjolfsson, 1993; Acemoglu, 2025).

For these reasons, short-term task improvements are interpreted in this review as evidence of productivity potential under defined conditions rather than as automatic evidence of durable productivity growth across individuals, organizations, or the economy as a whole.

5. Discussion

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 07/08/2026 - **Accepted:** 04/09/2026
Vol: 01.04
DOI: 10.66104/4hndmk68
Pages: 1-43

5.1 Conceptual Framework for AI-Assisted Productivity

Based on the evidence synthesized in this review, the productivity effects of Artificial Intelligence can be understood as the result of an interaction between technological capabilities, task characteristics, human expertise, verification requirements, and the organizational environment. AI capability alone is therefore insufficient to determine productivity outcomes. The value generated by an AI system depends on whether its capabilities are appropriate for the task being performed and whether users can effectively evaluate and integrate its outputs into existing workflows (Dell'Acqua et al., 2026; Peng et al., 2023; Ngo et al., 2025).

At the technological level, AI systems differ in their ability to generate content, retrieve and organize information, produce code, provide recommendations, and automate components of professional activities. These capabilities can support productivity through different mechanisms, including automation, assistance, augmentation, delegation, and human–AI collaboration. However, the effectiveness of these mechanisms depends on the characteristics of the task. Evidence reviewed in this study indicates that measurable benefits are more clearly demonstrated in selected tasks where objectives and outputs can be evaluated, whereas tasks involving ambiguous requirements, difficult-to-verify information, or outputs beyond the effective capability of the AI system may create greater risks and verification demands (Noy & Zhang, 2023; Peng et al., 2023; Dell'Acqua et al., 2026).

Human expertise constitutes another important component of this relationship. Less experienced users may obtain comparatively larger gains from AI-generated guidance in some tasks, while experienced users may retain important advantages in evaluating outputs, recognizing errors, and determining when AI-generated recommendations are inappropriate (Brynjolfsson et al., 2025; Dell'Acqua et al., 2026). Human expertise therefore influences both the benefits obtained from AI assistance and the costs associated with evaluating that assistance.

Verification represents a particularly important mechanism connecting AI assistance to realized productivity. An AI system may reduce the time or effort required to generate an initial output while simultaneously introducing additional work associated with checking factual accuracy, validating references, testing generated code, correcting errors, or ensuring compliance with professional requirements. Evidence concerning scientific references, software development, and other knowledge-intensive applications demonstrates that these verification requirements can be substantial when AI outputs are unreliable or difficult to evaluate (Cao, 2023; Peng et al., 2023; Ngo et al., 2025). Consequently, gross time savings should be

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 07/08/2026 - **Accepted:** 04/09/2026
Vol: 01.04
DOI: 10.66104/4hndmk68
Pages: 1-43

distinguished from net productivity gains.

At the organizational level, realized productivity also depends on conditions beyond the performance of an individual worker. Organizations must integrate AI systems into existing workflows, establish appropriate governance mechanisms, train users, manage data and security risks, and determine how human responsibility is maintained when tasks are partially automated. These complementary requirements help explain why improvements observed in individual tasks cannot automatically be extrapolated to equivalent improvements in organizational productivity (Ngo et al., 2025; Microsoft, 2024).

AI Capability → Task–AI Fit → AI Support Mechanism → Human Evaluation and Verification → Integrated Work Output → Productivity Outcome

The relationship is moderated by human expertise, verification cost, workflow integration, governance, and organizational readiness.

Net Productivity Effect = AI Assistance Benefit − Verification Cost − Integration Cost − Governance Cost

This expression is intended as a conceptual representation rather than a quantitative equation. Its purpose is to emphasize that observed improvements in generation speed or task execution represent only one component of the final productivity outcome. Where verification, correction, integration, or governance requirements are substantial, gross efficiency gains may not translate proportionally into net productivity improvements (Cao, 2023; Ngo et al., 2025).

The framework also distinguishes between levels of productivity. At the individual level, AI may influence task-completion time, output quantity, quality, and cognitive effort. At the team level, individual gains interact with coordination, review, knowledge sharing, and dependencies among workers. At the organizational level, productivity additionally depends on workflow redesign, training, governance, security, integration, and complementary investments. At the macroeconomic level, aggregate productivity depends on the diffusion of AI across economically relevant tasks, industries, and organizations and on whether technological change generates complementary forms of work or primarily substitutes existing activities (Brynjolfsson, 1993; Acemoglu, 2025).

Accordingly, the conceptual framework proposed in this review interprets AI-assisted productivity as a conditional and multilevel outcome rather than an automatic consequence of AI adoption. Productivity gains are more likely to emerge

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 07/08/2026 - **Accepted:** 04/09/2026
Vol: 01.04
DOI: 10.66104/4hndmk68
Pages: 1-43

when AI capabilities fit the requirements of the task, users possess sufficient expertise to evaluate outputs, verification costs remain manageable, and organizations provide the governance and integration mechanisms necessary to convert task-level efficiency into sustainable improvements. Conversely, poor task–AI fit, excessive reliance on unreliable outputs, high verification requirements, or inadequate organizational integration may reduce or offset the potential productivity benefits of AI assistance (Dell'Acqua et al., 2026; Ngo et al., 2025; Acemoglu, 2025).

Conceptual Framework for AI-Assisted Productivity

Synthesizing the empirical evidence and theoretical insights reviewed in this article, we propose a conceptual framework to understand how AI assistance affects productivity across tasks, individuals, organizations, and the economy (Figure 1).

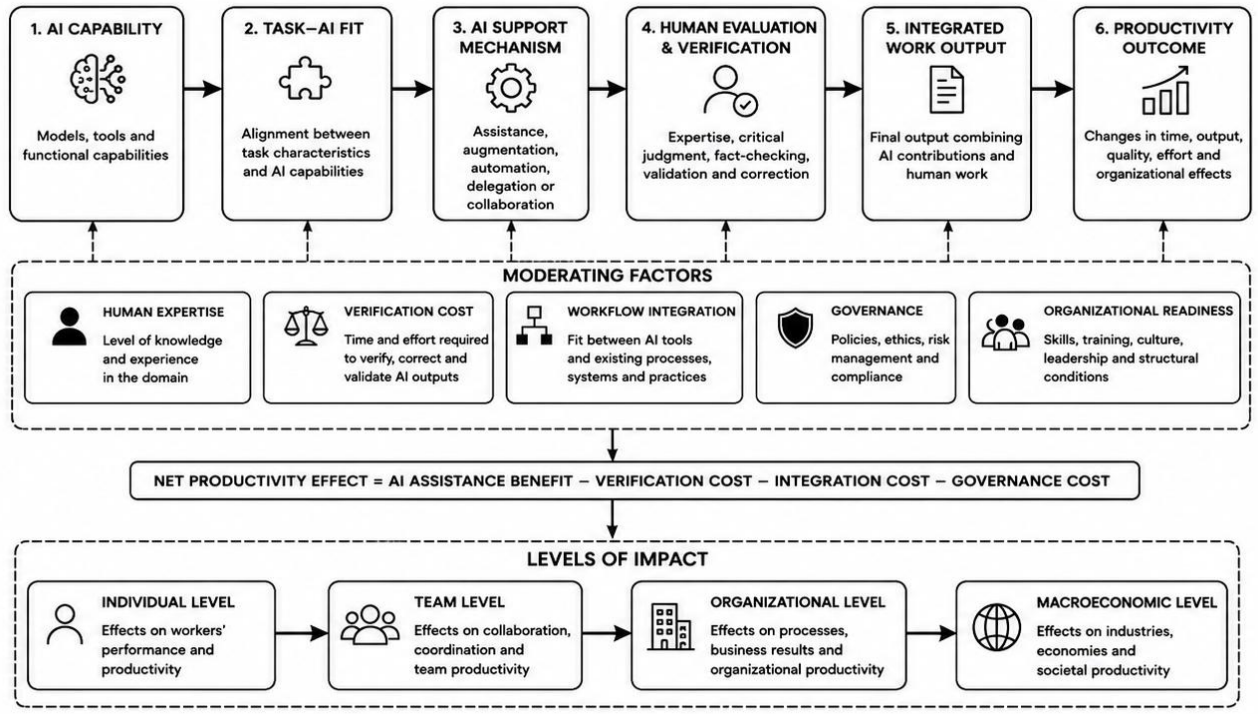

Figure 1. Conceptual Framework for AI-Assisted Productivity.

Note. The framework illustrates the sequential process through which AI capabilities influence productivity outcomes, while moderating factors shape the magnitude and direction of net productivity effects at multiple levels of analysis.

5.2 Future Perspectives

Future research on AI-assisted productivity is likely to examine how advances in model capability, multimodality, tool integration, and more autonomous forms of task execution affect the balance between assistance benefits and verification

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 07/08/2026 - **Accepted:** 04/09/2026
Vol: 01.04
DOI: 10.66104/4hndmk68
Pages: 1-43

requirements. Foundation-model research provides a technological basis for increasingly general-purpose systems, but evidence concerning their sustained productivity effects remains less mature than evidence for selected current tasks (Bommasani et al., 2021).

AI agents capable of coordinating multiple steps with reduced direct intervention represent an important research direction, but their productivity implications should be evaluated empirically rather than inferred from technical capability alone. Future studies should measure not only execution speed but also error propagation, supervision requirements, recovery from failures, and the human effort required to validate multi-step outputs (Bommasani et al., 2021; Ngo et al., 2025).

Multimodal systems may broaden the range of tasks that can be supported by combining text, image, audio, and other information. However, broader capability also creates new evaluation challenges, and future productivity research should distinguish technical capability from demonstrated net benefits in real workflows (Bommasani et al., 2021; Weidinger et al., 2022).

Nevertheless, future adoption will require continued attention to ethical governance. International frameworks, including UNESCO recommendations and OECD principles, emphasize the importance of transparency, accountability, safety, and human-centered AI development (UNESCO, 2021; OECD, 2019).

Overall, the reviewed literature suggests that future AI-assisted productivity will depend not only on technological progress but also on the ability of individuals and organizations to integrate these systems responsibly into human-centered workflows, with appropriate verification, governance, and accountability mechanisms (Ngo et al., 2025; UNESCO, 2021; OECD, 2019).

5.3 Limitations

This narrative and thematic literature review has several limitations that should be considered when interpreting its findings. First, the literature search was restricted to publications written in English and primarily focused on studies published between 2020 and 2026. Relevant evidence published in other languages or outside this period may therefore have been excluded. The searches used Google Scholar, IEEE Xplore, ACM Digital Library, ScienceDirect, SpringerLink, and Scopus. The original review record did not preserve an exact search date or date range, which limits reproducibility and should be addressed in future updates. Although these

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 07/08/2026 - **Accepted:** 04/09/2026
Vol: 01.04
DOI: 10.66104/4hndmk68
Pages: 1-43

sources provide broad multidisciplinary coverage, no combination of databases can guarantee complete retrieval of all relevant publications, and differences in database indexing may have influenced the literature identified for analysis.

Second, literature screening, eligibility assessment, data extraction, study classification, quality appraisal, and thematic coding were conducted by a single reviewer. Independent duplicate screening and coding were therefore not performed, and no inter-rater reliability statistic was calculated. Although predefined eligibility criteria and consistent analytical categories were used, the absence of an independent reviewer increases the possibility that individual judgment influenced study selection, classification, evidence appraisal, and thematic interpretation.

Third, the final analytical corpus consisted of 15 publications that contributed directly to the thematic synthesis, while 14 additional references served contextual, theoretical, ethical, historical, or institutional purposes. The relatively limited and heterogeneous analytical corpus restricts the extent to which the findings can be generalized to the rapidly expanding literature on AI-assisted productivity.

Fourth, although priority was given to peer-reviewed journal articles and scientific conference publications, selected preprints and working papers were retained when they addressed recent developments for which peer-reviewed evidence was limited. These publications were identified separately during evidence appraisal and were not treated as methodologically equivalent to peer-reviewed research. Nevertheless, their inclusion introduces additional uncertainty because their methods and conclusions may subsequently be revised following peer review.

Fifth, the included publications differ substantially in methodological design and evidential strength. The corpus contains controlled experiments, field studies, observational research, surveys, systematic reviews, meta-analyses, conference publications, and selected non-peer-reviewed studies. These designs differ in their ability to establish causal effects, their susceptibility to bias, their sample characteristics, and their external validity. Consequently, findings from different study types should not be interpreted as providing equivalent levels of evidence.

Sixth, productivity itself is operationalized differently across the reviewed domains. Some studies measure task-completion time or output volume, whereas others examine output quality, learning outcomes, worker performance, perceived usefulness, adoption, or reductions in cognitive and procedural effort. These outcomes are related but not interchangeable. In particular, faster task completion does not necessarily indicate a net productivity gain when additional verification,

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 07/08/2026 - **Accepted:** 04/09/2026
Vol: 01.04
DOI: 10.66104/4hndmk68
Pages: 1-43

correction, testing, integration, or governance effort is required.

Seventh, the evidence spans substantially different application domains, including professional knowledge work, software development, education, scientific research, and organizational activities. Productivity gains observed in one domain should not automatically be generalized to another because task structure, user expertise, error consequences, verification requirements, and acceptable levels of reliability differ across contexts. Similarly, task-level improvements observed in controlled or short-term settings cannot be assumed to translate proportionally into sustained team, organizational, or macroeconomic productivity.

Finally, Artificial Intelligence technologies are evolving considerably faster than conventional academic publication cycles. The capabilities, interfaces, reliability, and organizational integration of generative AI systems may change between the time a study is conducted and the time its results are published. Findings concerning a particular model or implementation may therefore become less representative as newer systems are introduced. This rapid technological evolution means that the present review should be interpreted as a synthesis of the evidence available within the defined period rather than as a definitive assessment of the long-term productivity effects of Artificial Intelligence.

Taken together, these limitations reinforce the need for caution when interpreting broad claims about AI-assisted productivity. The findings are most appropriately understood as evidence of heterogeneous and context-dependent effects rather than as evidence of a universal productivity advantage. Future research would benefit from longitudinal designs, standardized productivity measures, direct comparisons across levels of user expertise, clearer reporting of verification costs, and replication across different tasks, organizations, and AI systems.

6. Conclusion

This narrative and thematic literature review examined recent evidence concerning the use of Artificial Intelligence tools to support productivity across knowledge work, software development, education, scientific research, and organizational activities. The reviewed evidence suggests that generative AI can produce measurable productivity gains in selected tasks, particularly through reductions in task-completion time, support for information processing, content generation, coding assistance, and other forms of human–AI augmentation. However, these effects are heterogeneous and should not be interpreted as evidence of a universal productivity

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 07/08/2026 - **Accepted:** 04/09/2026
Vol: 01.04
DOI: 10.66104/4hndmk68
Pages: 1-43

improvement across users, tasks, organizations, or domains.

The synthesis indicates that task characteristics represent an important condition shaping AI-assisted productivity. Benefits are more clearly demonstrated in selected activities where the task is compatible with the capabilities of the AI system and where outputs can be evaluated against identifiable criteria. Conversely, tasks involving ambiguous requirements, difficult-to-verify information, or outputs outside the effective capabilities of the AI system may produce smaller benefits or greater risks. The possibility and cost of verifying AI-generated outputs are therefore central to determining whether initial time savings translate into net productivity gains (Cao, 2023; Peng et al., 2023; Dell'Acqua et al., 2026).

User expertise also influences productivity outcomes. Evidence indicates that less experienced users may obtain comparatively larger gains from AI-generated guidance in some settings, while expertise remains important for identifying errors, evaluating relevance, and deciding when AI-generated outputs should be accepted, modified, or rejected (Noy & Zhang, 2023; Brynjolfsson et al., 2025; Dell'Acqua et al., 2026). AI assistance should therefore be understood primarily as a context-dependent mechanism for automation, assistance, augmentation, delegation, or collaboration rather than as a uniform substitute for human expertise.

At the organizational level, the realization of productivity benefits depends on more than the technical capabilities of AI systems. Workflow integration, employee training, human oversight, data protection, governance mechanisms, and organizational readiness can influence whether task-level efficiency gains become sustainable improvements in organizational performance (Ngo et al., 2025; Microsoft, 2024). Verification, correction, integration, and governance requirements should consequently be considered part of the total cost of AI-assisted productivity rather than external issues unrelated to productivity.

The evidence also varies considerably in methodological strength. Controlled experiments and field studies provide comparatively direct evidence of productivity effects for specific tasks, while surveys, adoption studies, evidence syntheses, preprints, and working papers address different questions and provide different levels of evidential certainty. In addition, productivity is operationalized through heterogeneous measures, including completion time, output quantity, quality, learning outcomes, cognitive effort, and organizational performance. These differences limit direct comparison and reinforce the need for evidence-bounded conclusions.

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 07/08/2026 - **Accepted:** 04/09/2026
Vol: 01.04
DOI: 10.66104/4hndmk68
Pages: 1-43

Overall, the reviewed literature supports a conditional interpretation of AI-assisted productivity. AI tools can generate meaningful productivity benefits when their capabilities are appropriately matched to task requirements, users possess sufficient expertise to evaluate their outputs, verification costs remain manageable, and organizational conditions support effective and responsible integration. Conversely, poor task–AI fit, unreliable outputs, high verification requirements, inadequate training, or weak governance may reduce or offset these benefits. Task-level gains should therefore not automatically be extrapolated to long-term organizational or macroeconomic productivity (Brynjolfsson, 1993; Acemoglu, 2025).

Future research should prioritize longitudinal and comparative studies capable of determining whether short-term performance improvements persist over time and translate into broader organizational outcomes. Greater standardization of productivity measures would facilitate comparison across studies, while explicit measurement of verification and correction costs would provide a more realistic estimate of net productivity effects. Research should also examine differences across user expertise levels, task types, organizational environments, and successive generations of AI systems. As Artificial Intelligence continues to evolve rapidly, evaluating not simply whether AI makes individual tasks faster, but under which conditions those improvements become reliable, sustainable, and transferable productivity gains will remain a central research challenge.

Funding

The author received no financial support for the research, authorship, or publication of this article.

Conflicts of Interest

The author declares that there are no conflicts of interest regarding the publication of this manuscript.

Author Responsibility Statement

The author confirms substantial responsibility for the conception, design, execution, interpretation, and final approval of this manuscript and agrees to its submission to the journal.

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 07/08/2026 - **Accepted:** 04/09/2026
Vol: 01.04
DOI: 10.66104/4hndmk68
Pages: 1-43

References

Acemoglu, D. (2025). The simple macroeconomics of AI. Economic Policy, 40(121), 13–58. https://doi.org/10.1093/epolic/eiae042

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (pp. 610–623). Association for Computing Machinery. https://doi.org/10.1145/3442188.3445922

Bick, A., Blandin, A., & Deming, D. J. (2026). The rapid adoption of generative AI. Management Science. https://doi.org/10.1287/mnsc.2025.02523

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., ... & Liang, P. (2021). On the opportunities and risks of foundation models. arXiv. https://doi.org/10.48550/arXiv.2108.07258

Brynjolfsson, E. (1993). The productivity paradox of information technology. Communications of the ACM, 36(12), 66–77. https://doi.org/10.1145/163298.163309

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

Cao, Z. (2023). Evaluating the efficacy of ChatGPT-4 in providing scientific references across diverse disciplines. arXiv. https://doi.org/10.48550/arXiv.2306.09914

Clark, J., Barton, B. L., Albarqouni, L., Byambasuren, O., Jowsey, T., Keogh, J. W. L., Liang, T., Moro, C., O'Neill, H. M., & Jones, M. A. (2025). Generative artificial intelligence use in evidence synthesis: A systematic review. Research Synthesis Methods, 16(4), 601–619. https://doi.org/10.1017/rsm.2025.16

de la Torre-López, J., Ramírez, A., & Romero, J. R. (2023). Artificial intelligence to automate the systematic review of scientific literature. Computing, 105, 2171–2194. https://doi.org/10.1007/s00607-023-01181-x

Dell'Acqua, F., McFowland III, 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.

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 07/08/2026 - **Accepted:** 04/09/2026
Vol: 01.04
DOI: 10.66104/4hndmk68
Pages: 1-43

Organization Science, 37(2), 403–423. https://doi.org/10.1287/orsc.2025.21838

Ebrahim, S., & Tanner, M. (2025). A systematic literature review on the use of ChatGPT to support productivity of software developers. In Proceedings of the 31st Americas Conference on Information Systems (AMCIS 2025).

Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. Harvard Data Science Review, 1(1). https://doi.org/10.1162/99608f92.8cd550d1

Gabashvili, I. S. (2023). The impact and applications of ChatGPT: A systematic review of literature reviews. arXiv. https://doi.org/10.48550/arXiv.2311.04570

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press. https://www.deeplearningbook.org/

Humlum, A., & Vestergaard, E. (2025). The adoption of ChatGPT. University of Chicago, Becker Friedman Institute Working Paper. https://bfi.uchicago.edu/working-paper/the-adoption-of-chatgpt/

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. Science, 349(6245), 255–260. https://doi.org/10.1126/science.aaa8415

Kaplan, A., & Haenlein, M. (2020). Rulers of the world, unite! The challenges and opportunities of artificial intelligence. Business Horizons, 63(1), 37–50. https://doi.org/10.1016/j.bushor.2019.09.003

Maier, S., Gunzenhäuser, M., Schweisthal, J., Schneider, M., & Feuerriegel, S. (2026). A meta-analysis of the effect of generative AI on productivity and learning in programming. arXiv. https://doi.org/10.48550/arXiv.2605.04779

Microsoft. (2024). 2024 Work Trend Index Annual Report: AI at work is here. Now comes the hard part. Microsoft. https://www.microsoft.com/worklab/work-trend-index

Ngo, C.-L., Soyoof, A., & Tseringa, D. (2025). A systematic review of the limitations and associated opportunities of ChatGPT. International Journal of Human–Computer Interaction, 41(7), 3851–3866. https://doi.org/10.1080/10447318.2024.2344142

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of

CC BY

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 07/08/2026 - **Accepted:** 04/09/2026
Vol: 01.04
DOI: 10.66104/4hndmk68
Pages: 1-43

generative artificial intelligence. Science, 381(6654), 187–192. https://doi.org/10.1126/science.adh2586

Organisation for Economic Co-operation and Development. (2019). OECD principles on artificial intelligence. https://oecd.ai/en/ai-principles

Peng, S., Kalliamvakou, E., Cihon, P., & Demirer, M. (2023). The impact of AI on developer productivity: Evidence from GitHub Copilot. arXiv. https://doi.org/10.48550/arXiv.2302.06590

Rahman, M. M., Al Mamun, M. A., Hasan, M. K., Islam, M. R., Hossain, M. S., & Rahman, M. M. (2025). The influence of artificial intelligence tools on learning outcomes in computer programming: A systematic review and meta-analysis. Computers, 14(5), Article 185. https://doi.org/10.3390/computers14050185

Russell, S. J., & Norvig, P. (2021). Artificial intelligence: A modern approach (4th ed.). Pearson.

Singh, H. V. (2026). Generative AI and worker productivity: A systematic review and quantitative evidence synthesis (2023–2026). SSRN. https://doi.org/10.2139/ssrn.6366218

Teubner, T., Flath, C. M., Weinhardt, C., van der Aalst, W. M. P., & Hinz, O. (2023). Welcome to the era of ChatGPT: The prospects of large language models. Business & Information Systems Engineering, 65(2), 95–101. https://doi.org/10.1007/s12599-023-00793-0

United Nations Educational, Scientific and Cultural Organization. (2021). Recommendation on the ethics of artificial intelligence. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000381137

Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P.-S., Cheng, M., Glaese, A., Balle, B., Kasirzadeh, A., Kenton, Z., Brown, S., Hawkins, W., Stepleton, T., Birhane, A., Haas, J., Rimell, L., Hendricks, L. A., Isaac, W., ... Gabriel, I. (2022). Taxonomy of risks posed by language models. In Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (pp. 214–229). Association for Computing Machinery. https://doi.org/10.1145/3531146.3533088

CC BY