SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

Intelligent Business Automation through Agentic AI: Enhancing Workflow Efficiency and Adaptive Decision-Making

Heta Patel

Sr SAP Developer Doterra International LLC Saratoga Springs, Utah, USA. E-mail: heta4172@gmail.com

Abstract
Artificial intelligence is increasingly being integrated into business environments to automate routine activities, support decision-making, and improve workflow management. Agentic AI extends conventional automation by enabling intelligent systems to interpret objectives, reason about tasks, and perform actions within workplace environments. This study aimed to characterize workplace automation requirements by examining the distribution of tasks across functional domains and assessing the workflow actions required for task completion. Workplace automation tasks from the WorkBench benchmark were analyzed across six functional domains, including analytics, calendar management, customer relationship management, email management, multi-domain workflows, and project management. Tasks were examined according to their workflow-action requirements. Descriptive analysis was performed using task frequencies, percentages, and domain-level comparisons to characterize workplace automation requirements. The findings demonstrated that workplace automation tasks were distributed across multiple functional domains, with multi-domain workflows representing an important component of the benchmark. Most tasks required at least one workflow action, although the level of action requirement varied across functional areas. Email management showed a particularly high proportion of action-required tasks, whereas analytics contained a comparatively larger proportion of tasks requiring no direct workflow action. These findings demonstrate substantial variation in workplace task structures. Workplace automation involves diverse functional requirements and different levels of workflow interaction. The findings highlight the importance of developing flexible agentic AI systems capable of adapting task execution to different workplace contexts and supporting intelligent workflow management and adaptive decision-making.

Keywords: Artificial intelligence, Business process automation, Decision support systems, Intelligent automation, Workflow optimization

1. Introduction

Artificial intelligence (AI) has become an important technology for improving how organizations manage information, coordinate activities, and automate repetitive operations. Earlier business automation approaches generally relied on predefined rules, fixed workflows, and structured inputs. Such systems can perform repetitive activities consistently, but their ability to respond to changing information and unstructured instructions is limited. The development of large language models (LLMs) has expanded the capabilities of AI by enabling systems to interpret natural-language instructions and generate context-sensitive responses (Kodner *et al*., 2022). Further developments in language-model reasoning have demonstrated the ability of these systems to address increasingly complex tasks through structured reasoning processes (Wei *et al*., 2022). These developments provide a foundation for business automation systems capable of supporting activities that require interpretation, reasoning, and interaction rather than simple rule execution.

The evolution of AI has progressed from conventional rule-based automation toward systems capable of planning and performing actions with greater autonomy. Traditional automation generally follows predetermined sequences, whereas agentic AI systems can interpret objectives, determine intermediate steps, and interact with external tools or environments (Fettke and Di Francescomarino, 2025). The concept of autonomous LLM-based agents has received substantial research attention; these systems combine language understanding with planning, reasoning, memory, and action capabilities (Wang *et al*., 2024). A broader review of LLM-based agents further describes their potential for autonomous task completion and interaction with external environments. Agent-based approaches therefore provide a pathway toward automation systems that can adapt their behavior according to task requirements rather than simply following fixed procedural instructions (Ghane *et al*., 2026).

Agentic AI is particularly relevant to autonomous decision-making; an agent can use information from its environment to determine an appropriate sequence of actions (Priyadarshi, 2025). Reasoning-based approaches have demonstrated how language models can improve their ability to solve multistep problems by explicitly developing intermediate reasoning processes (Yao et al., 2023). Tool-augmented language models extend this capability by allowing models to interact with application programming interfaces and other external functions (Li *et al*., 2023). Large-scale API integration has further demonstrated the potential for language models to select and operate external tools for completing tasks (Patil *et al*., 2024). Agent systems can also incorporate feedback from previous attempts to improve subsequent decisions, as demonstrated by Reflexion (Shinn *et al*., 2023). These capabilities are important for intelligent business automation; workplace tasks may require an agent to interpret an instruction, select an action, evaluate the resulting state, and continue until the specified objective is achieved (Hughes *et al*., 2025).

Business-process management represents an important application area for intelligent automation; organizational workflows frequently contain interconnected activities, information sources, decisions, and software interactions. Generative AI has consequently been investigated as a means of transforming business-process management and supporting process-related activities (Beheshti *et al*., 2023). Business-process-aware language models have also been proposed to facilitate interaction between users, language models, and process information (Bernardi *et al*., 2024). Research examining BPM in the age of AI indicates that artificial intelligence is influencing how business processes are designed, managed, and improved (Rosemann *et al*., 2024). AI-supported BPM therefore provides an important context for agentic systems; autonomous agents can potentially connect natural-language objectives with the execution of multiple workflow activities (Janakiraman, 2026). This creates opportunities for automation that extends beyond individual tasks toward coordinated business processes.

Despite these developments, deploying autonomous AI systems for workplace automation involves several challenges. Agents must correctly interpret user requirements, select appropriate tools, maintain task context, and perform actions without introducing unintended changes (Jaggavarapu, 2025). Benchmark research has therefore emphasized the importance of evaluating autonomous agents in realistic environments and across diverse tasks. Web-based environments have similarly been developed to assess whether agents can perform complex tasks involving multiple websites and interactions (Zhou *et al*., 2024). Research on multimodal computer-use agents has further demonstrated the difficulty of performing open-ended tasks in realistic computer environments (Xie *et al*., 2024). Responsible deployment also requires attention to transparency, reliability, accountability, and human oversight. Responsible AI research in BPM highlights the importance of considering these factors when AI is used to improve organizational processes (Pisoni and Moloney, 2024).

Although agentic AI research has expanded rapidly, important gaps remain in understanding how autonomous AI capabilities correspond to workplace workflow requirements. Existing studies have investigated web interaction, tool use, software engineering, reasoning, and general autonomous agents. For example, Mind2Web examined generalist agents performing tasks on real-world websites (Deng *et al*., 2023). SWE-agent demonstrated the use of agent-computer interfaces for autonomous software-engineering tasks (Yang *et al*., 2024). WebShop investigated grounded language agents interacting with web-based environments to accomplish practical objectives (Yao et al., 2022). These studies demonstrate substantial progress in autonomous task execution, but they do not by themselves provide a complete characterization of workplace automation requirements across business functions (Parab, 2024). A systematic examination of workplace tasks can therefore contribute to understanding the functional distribution and action requirements that intelligent business automation systems need to address (Park *et al*., 2023).

The study aims to characterize workplace automation requirements within an agentic-AI framework by examining workplace tasks across multiple functional domains and identifying the workflow actions required for their completion. Specifically, the study has two objectives: first, to assess the distribution of workplace automation tasks across functional domains and determine the proportion of tasks requiring workflow actions; and second, to examine workflow action requirements across functional domains to characterize task complexity relevant to intelligent business automation and adaptive decision-making. This focus is consistent with emerging research examining autonomous AI systems and their application to business-oriented decision-making and operational processes.

2. Materials and Methods

2.1 Study Design

This study used the publicly available WorkBench benchmark to examine workplace automation tasks and their workflow action requirements. WorkBench was developed for evaluating agents performing workplace activities through tool-based interactions. The benchmark contains 690 tasks covering analytics, calendar management, customer relationship management, email management, project management, and multi-domain workflows. Each task includes a natural-language instruction and a predefined ground-truth outcome describing the required final action state. The benchmark uses a controlled workplace environment containing five databases and 26 tools. Tasks are evaluated by comparing the resulting environment state with the predefined ground truth, allowing consistent assessment of workflow requirements.

2.2 Dataset Source

The dataset was obtained from the publicly released WorkBench repository and consisted of the corrected task-and-outcome files used for the current benchmark version. The six task files represented analytics, calendar management, customer relationship management, email management, multi-domain workflows, and project management. Their respective task counts were 120, 110, 80, 90, 210, and 80, giving a total of 690 tasks. The benchmark also provides five workplace-oriented databases containing analytics records, calendar events, customer relationship records, emails, and project-management tasks. These resources provide structured workplace information for evaluating agent-oriented task execution and workflow requirements within a controlled environment.

2.3 Data Variables

Each task-and-outcome file contained five principal fields: task, outcome, base template, chosen template, and domains. The task field represented the natural-language workplace request, whereas the outcome field specified the ground-truth action sequence associated with that request. The base-template and chosen-template fields described the template structures used to generate individual tasks. The domains field identified the functional area or combination of areas involved. For the present analysis, functional domain and outcome information were the principal variables. The number of required workflow actions was determined from the actions listed within each ground-truth outcome, allowing tasks to be grouped according to their operational action requirements.

2.4 Data Processing

The six task-and-outcome files were combined to obtain the complete set of 690 workplace tasks. Functional-domain labels were retained according to the dataset files, while the outcome field was examined to determine the number of predefined actions associated with each task. An empty outcome list was treated as requiring no action. Non-empty outcome lists were counted according to the number of actions they contained. Frequencies were calculated for each action-count category, and percentages were obtained by dividing each frequency by the total number of tasks and multiplying by 100. Domain-specific action requirements were summarized using task counts, no-action frequencies, action-required frequencies, and mean required actions per task.

2.5 Statistical Analysis

Descriptive statistical procedures were used to summarize the distribution of workplace automation tasks and their workflow requirements. Categorical variables were presented as frequencies and percentages. The mean number of required actions was calculated for each functional domain by dividing the total number of predefined actions by the corresponding number of tasks. Overall task distributions were also summarized according to the number of required actions. Results were organized into three tables describing functional-domain distribution, action-count distribution, and domain-specific workflow requirements. Corresponding bar figures were prepared to visually represent the principal distributions.

2.6 Data Validation

The analysis was based exclusively on the task-and-outcome files contained in the WorkBench repository supplied for this study. The benchmark documentation describes 690 workplace tasks, five sandbox databases, and 26 tools, with task performance evaluated against predefined ground-truth final states. The WorkBench release also identifies corrections to ground-truth and task-design issues in the updated benchmark.

3. Results

3.1 Distribution of Workplace Automation Tasks by Functional Domain

The WorkBench dataset comprised 690 workplace automation tasks distributed across six functional domains. Multi-domain workflows represented the largest category, with 210 tasks (30.43%), followed by analytics with 120 tasks (17.39%). Calendar management accounted for 110 tasks (15.94%), while email management included 90 tasks (13.04%), as shown in Table 1. Customer relationship management and project management each contributed 80 tasks, representing 11.59% of the dataset. The distribution demonstrates that the benchmark covers diverse workplace activities rather than concentrating on a single operational function. This broad domain coverage provides a structured basis for examining agentic AI applications across different business workflow requirements and automation contexts.

Table 1. Distribution of Workplace Automation Tasks by Functional Domain

Functional domain	Number of tasks	Percentage (%)
Analytics	120	17.39
Calendar management	110	15.94
Customer relationship management	80	11.59
Email management	90	13.04
Multi-domain workflows	210	30.43
Project management	80	11.59
Total	690	100.00

Source: *WorkBench 2026 corrected task-and-outcome dataset.*

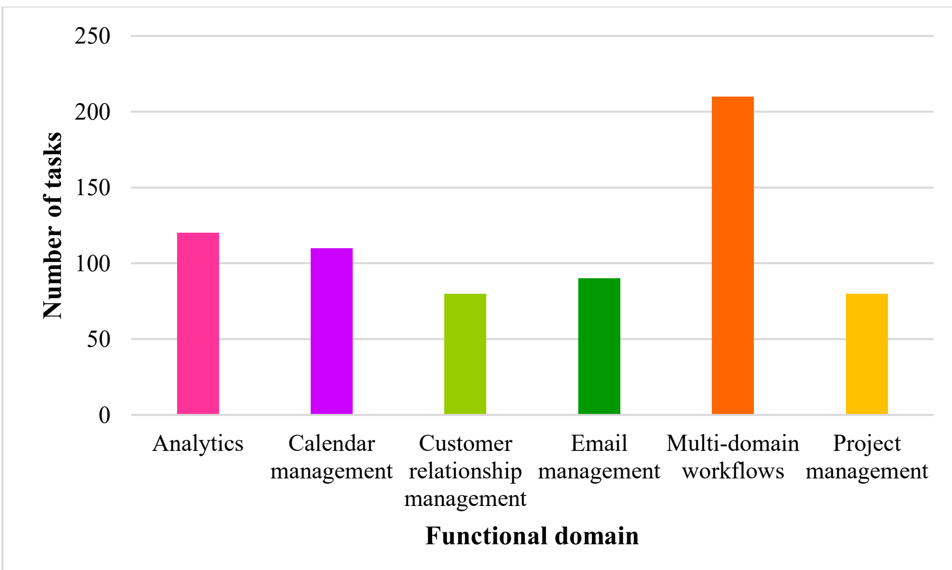

Figure 1. Distribution of Workplace Automation Tasks by Functional Domain

The distribution of 690 workplace automation tasks across six functional domains. Multi-domain workflows represented the largest category, with 210 tasks. Analytics accounted for 120 tasks, followed by calendar management with 110 tasks, as shown in Figure 1. Email management included 90 tasks, while customer relationship management and project management each comprised 80 tasks. The distribution demonstrates that the dataset covers several workplace functions rather than focusing on one operational area. Multi-domain workflows therefore represented the largest share of tasks, whereas customer relationship management and project management had comparatively smaller task counts. Overall, the figure shows broad functional coverage within the WorkBench dataset.

3.2 Distribution of Workplace Tasks According to Required Workflow Actions

Across the 690 workplace tasks, most tasks required at least one predefined workflow action. A total of 416 tasks (60.29%) required one action, whereas 88 tasks (12.75%) required two actions. Smaller proportions required three actions (3.33%) or four actions (2.03%), as shown in Table 2. Tasks involving five or more actions were relatively uncommon, with individual categories ranging from 0.29% to 0.87%. In contrast, 122 tasks (17.68%) required no action according to their ground-truth outcomes. The distribution indicates considerable variation in workflow requirements, ranging from tasks requiring no intervention to tasks involving multiple sequential actions within the workplace environment.

Table 2. Distribution of Workplace Tasks According to Required Workflow Actions

Required actions	Number of tasks	Percentage (%)
0	122	17.68
1	416	60.29
2	88	12.75
3	23	3.33
4	14	2.03
5	2	0.29
6	5	0.72
7	3	0.43
8	4	0.58
9	3	0.43
10	2	0.29
11	6	0.87
12	2	0.29
Total	690	100.00

Source: *WorkBench 2026 corrected task-and-outcome dataset.*

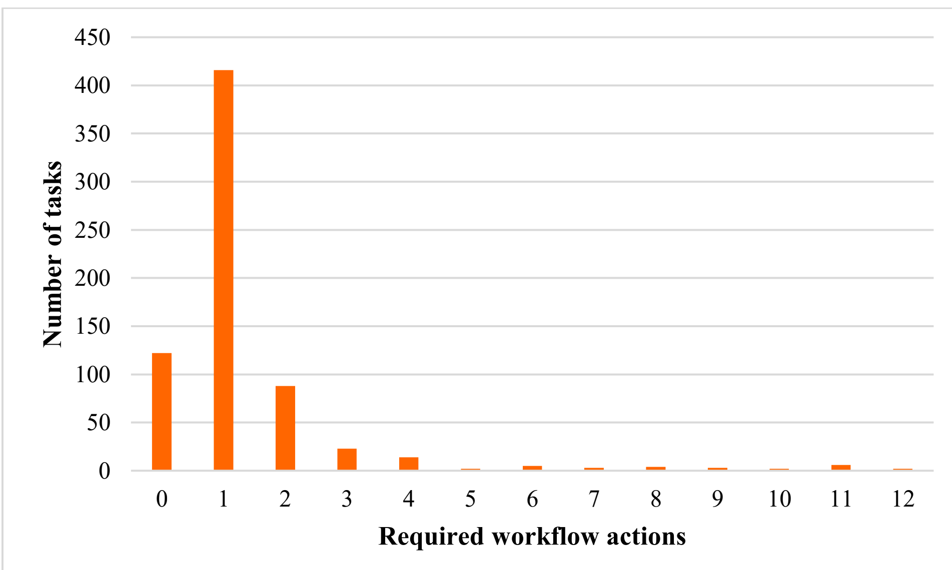

Figure 2. Distribution of Workplace Tasks by Workflow Action Requirements

The distribution of workplace tasks according to the number of required workflow actions. Tasks requiring one action were most frequent, followed by tasks requiring two actions. The number of tasks decreased substantially as the required action count increased beyond two, as shown in Figure 2. Tasks requiring three actions were less frequent, while tasks requiring four or more actions represented relatively small proportions of the dataset. The pattern indicates that most workplace tasks involve a limited number of predefined actions, whereas highly action-intensive workflows occur less frequently. This distribution provides an overview of workflow complexity within the WorkBench benchmark and its workplace automation tasks.

3.3 Workflow Action Requirements Across Functional Domains

Workflow action requirements varied across the six functional domains. Multi-domain workflows demonstrated the highest mean requirement, with 1.78 actions per task, followed by customer relationship management at 1.58. Project management averaged 1.26 actions, while email and calendar management averaged 1.22 and 1.19 actions, respectively, as shown in Table 3. Analytics showed the lowest mean at 0.88 actions per task. Overall, 568 tasks (82.32%) required at least one action, compared with 122 tasks (17.68%) requiring no action. These findings indicate that workplace automation tasks differ substantially in operational complexity, with multi-domain and customer relationship activities generally involving greater action requirements within the benchmark.

Table 3. Workflow Action Requirements Across Functional Domains

Functional domain	Total tasks	No-action tasks, n (%)	Action-required tasks, n (%)	Mean required actions
Analytics	120	40 (33.33)	80 (66.67)	0.88
Calendar management	110	11 (10.00)	99 (90.00)	1.19
Customer relationship management	80	8 (10.00)	72 (90.00)	1.58
Email management	90	1 (1.11)	89 (98.89)	1.22
Multi-domain workflows	210	41 (19.52)	169 (80.48)	1.78
Project management	80	21 (26.25)	59 (73.75)	1.26
Overall	690	122 (17.68)	568 (82.32)	1.37

Source: *WorkBench 2026 corrected task-and-outcome dataset.*

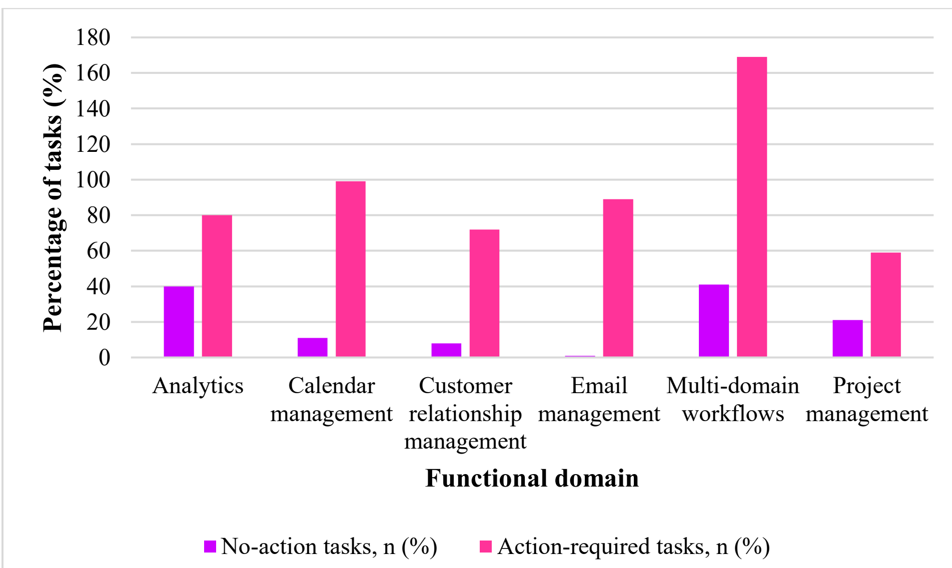

Figure 3. Workflow Action Requirements Across Functional Domains

The proportions of workplace tasks requiring actions versus those requiring no action across six functional domains. Email management showed the highest proportion of action-required tasks at 98.89%, followed by calendar management and customer relationship management, both at 90.00%, as shown in Figure 3. Multi-domain workflows recorded 80.48% action-required tasks, while project management accounted for 73.75%. Analytics had the lowest proportion of action-required tasks at 66.67% and the highest proportion of no-action tasks at 33.33%. Overall, 82.32% of tasks required at least one workflow action, whereas 17.68% required no action, indicating variation in workflow requirements across functional domains.

4. Discussion

The study examined workplace automation requirements across 690 tasks within the WorkBench benchmark, focusing on functional distribution and the number of workflow actions required for task completion. The findings demonstrate that workplace automation encompasses multiple operational domains rather than a single type of activity. Multi-domain workflows represented the largest category, while analytics, calendar management, email management, customer relationship management, and project management also contributed substantial numbers of tasks. Overall, 82.32% of tasks required at least one workflow action, whereas 17.68% required no action. These findings indicate that workplace automation frequently involves interaction with operational environments rather than simple information generation.

The distribution of tasks across functional domains demonstrates the heterogeneous nature of workplace automation. Multi-domain workflows accounted for 210 tasks, representing the largest category, which indicates that many workplace activities may involve interactions across more than one functional area. Analytics accounted for 120 tasks, followed by calendar management with 110 tasks and email management with 90 tasks. Customer relationship management and project management each contributed 80 tasks. This distribution provides evidence that an intelligent workplace agent may need to operate across different types of information and applications rather than being designed for one narrowly defined business function. The breadth of environments used in autonomous-agent research similarly demonstrates the importance of evaluating agents across diverse task settings rather than relying on a single interaction environment.

The workflow-action analysis provides an important indication of task complexity. Most tasks required a relatively small number of actions, with one-action tasks representing the largest group at 416 tasks. Tasks requiring two actions accounted for 88 cases, while increasingly action-intensive workflows occurred less frequently. At the same time, the presence of tasks requiring several actions demonstrates that workplace automation cannot be understood solely as single-step command execution. An agent may need to determine a sequence of operations and maintain an appropriate task state while completing the requested objective. Deliberate reasoning approaches have demonstrated the relevance of structured intermediate reasoning when language models address problems involving multiple possible steps (Yao *et al*., 2023). Thus, workflow action requirements provide a useful perspective for examining the operational demands placed on agentic systems.

The proportion of action-required tasks differed considerably among functional domains. Email management had the highest proportion of action-required tasks at 98.89%, followed by calendar management and customer relationship management, both at 90.00%. Multi-domain workflows showed 80.48% action-required tasks, while project management recorded 73.75%. Analytics had the lowest proportion of action-required tasks at 66.67%. These differences suggest that functional context can influence the type of interaction expected from an automation agent. Domains involving communication, scheduling, or record modification may require more direct system interaction, whereas some analytics-oriented tasks may be completed through information retrieval or interpretation. Similar differences in task interaction requirements have been observed in environments designed to evaluate agents performing real computer-based activities (Xi *et al*., 2025).

The findings have implications for the design of agentic AI systems intended for workplace automation. Tasks vary in both functional domain and action requirements; a general-purpose workplace agent should be capable of interpreting different task objectives and selecting appropriate actions. The importance of connecting agents with external interfaces is demonstrated by systems designed for grounded interaction with web environments and applications (Yao *et al*., 2022). The present findings extend this perspective to workplace-oriented tasks by showing that automation requirements can vary substantially across functional domains. Consequently, agent architecture should accommodate different levels of workflow interaction rather than assuming that all business tasks follow an identical execution pattern.

Adaptive decision-making is particularly important when a task requires multiple actions or when the appropriate next operation depends on the current state of an environment. A workplace agent that simply executes a fixed sequence may be less flexible when intermediate outcomes influence subsequent actions. Research on autonomous agents has demonstrated the importance of combining reasoning with interaction and task execution. For example, WebArena evaluates agents in realistic web environments where successful completion requires interaction with multiple components and services. Similarly, SWE-agent demonstrates how agent-computer interfaces can support autonomous execution of complex tasks within software environments. These findings support the relevance of adaptive workflow management for workplace-oriented agentic AI.

From an organizational perspective, the findings suggest that agentic automation should be considered according to the characteristics of individual workflows rather than treated as a uniform technological intervention. Tasks involving direct actions, multiple systems, or sequential operations may require stronger tool-use and state-management capabilities. In contrast, tasks requiring limited interaction may be suitable for simpler forms of AI assistance. This distinction can help organizations identify where autonomous agents may need greater supervision, tool access, or workflow controls. The results also provide a structured basis for evaluating workplace automation requirements before deploying agentic systems across different operational functions.

Several limitations should be considered when interpreting these findings. First, the analysis describes the distribution and action requirements of tasks rather than measuring the performance of a specific AI agent. Therefore, the results should not be interpreted as evidence that an agent can complete all identified workflows. Second, the WorkBench environment represents a controlled workplace-agent benchmark rather than direct observational data from an individual company's live operational systems. Third, task complexity was characterized primarily through workflow-action requirements and therefore does not capture every dimension of complexity, such as reasoning difficulty, ambiguity, execution risk, or organizational constraints. Real-world computer-use research similarly demonstrates that autonomous agents face substantial challenges when operating in open-ended environments.

Future studies can build on these findings by evaluating how different agent architectures perform across the functional domains and workflow-action levels identified in the present study. Studies could compare autonomous agents according to task-completion accuracy, action-selection reliability, execution efficiency, error rates, and human intervention requirements. More complex investigations could also examine how agents respond when intermediate workflow states change unexpectedly. Web-based agent research demonstrates the value of realistic environments for assessing autonomous task completion. Future studies should therefore combine task-level characterization with systematic agent-performance evaluation to determine how effectively agentic AI can translate workplace objectives into reliable and adaptive workflow execution.

5. Conclusion

The study characterized 690 workplace automation tasks across six functional domains and identified substantial variation in their workflow requirements. Multi-domain workflows represented the largest task category, while email, calendar, analytics, customer relationship management, and project management also contributed meaningful task volumes. Overall, most tasks required at least one workflow action, demonstrating that workplace automation frequently involves interaction with operational environments rather than information generation alone. The findings indicate that agentic AI systems should be capable of handling different workplace functions and varying levels of workflow interaction. Systems designed for business automation may require flexible action selection, task-state management, and the ability to coordinate multiple operations according to the requirements of individual tasks. This study provides a structured characterization of workplace automation requirements based on functional domains and workflow actions. Such characterization can support the development and evaluation of agentic systems intended to perform workplace tasks and assist organizations in understanding the operational requirements of intelligent automation. The analysis was based on a controlled workplace-agent benchmark rather than live organizational operational records. In addition, workflow-action counts do not capture every dimension of task difficulty, such as reasoning complexity or execution risk. Future research should evaluate agent performance across these task categories using measures such as completion accuracy, action reliability, efficiency, and human intervention. Further studies could also examine adaptive behavior in more dynamic workplace environments.

References

1. Beheshti, A., Yang, J., Sheng, Q. Z., Benatallah, B., Casati, F., Dustdar, S., ... & Xue, S. (2023, July). ProcessGPT: transforming business process management with generative artificial intelligence. In *2023 IEEE international conference on web services (ICWS)* (pp. 731-739). IEEE.
2. Bernardi, M. L., Casciani, A., Cimitile, M., & Marrella, A. (2024). Conversing with business process-aware large language models: the BPLLM framework. *Journal of Intelligent Information Systems*, *62*(6), 1607-1629.
3. Deng, X., Gu, Y., Zheng, B., Chen, S., Stevens, S., Wang, B., ... & Su, Y. (2023). Mind2web: Towards a generalist agent for the web. *Advances in Neural Information Processing Systems*, *36*, 28091-28114.
4. Fettke, P., & Di Francescomarino, C. (2025). Business process management and artificial intelligence: Literature survey and future research. *KI-Künstliche Intelligenz*, *39*(2), 67-79.

5. Ghane, M., Sorooshian, S., & Ang, M. C. (2026). Artificial Intelligence agents and autonomous decision-making in business: a review. *Cogent Business & Management*, *13*(1), 2692205.
6. Hughes, L., Dwivedi, Y. K., Malik, T., Shawosh, M., Albashrawi, M. A., Jeon, I., ... & Walton, P. (2025). AI agents and agentic systems: A multi-expert analysis. *Journal of Computer Information Systems*, *65*(4), 489-517.
7. Jaggavarapu, M. K. R. (2025). The evolution of agentic AI: architecture and workflows for autonomous systems. *Journal Of Multidisciplinary*, *5*(7), 418-427.
8. Janakiraman, A. (2026). Agentic Large Language Models for Autonomous Decision-Making and Adaptive Task Orchestration in Intelligent Systems. *International Journal of Sustainable Digital and Computing Systems*, *3*(1).
9. Kodner, J., Caplan, S., & Yang, C. (2022). Another model not for the learning of language. *Proceedings of the National Academy of Sciences*, *119*(29), e2204664119.
10. Li, M., Zhao, Y., Yu, B., Song, F., Li, H., Yu, H., ... & Li, Y. (2023, December). Api-bank: A comprehensive benchmark for tool-augmented llms. In *Proceedings of the 2023 conference on empirical methods in natural language processing* (pp. 3102-3116).
11. Parab, G. U. (2024). Agentic AI in data analytics: Transforming autonomous insights and decision-making. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, *10*(6), 1752-1759.
12. Park, J. S., O'Brien, J., Cai, C. J., Morris, M. R., Liang, P., & Bernstein, M. S. (2023, October). Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th annual acm symposium on user interface software and technology* (pp. 1-22).
13. Patil, S. G., Zhang, T., Wang, X., & Gonzalez, J. E. (2024). Gorilla: Large language model connected with massive apis. *Advances in Neural Information Processing Systems*, *37*, 126544-126565.
14. Pisoni, G., & Moloney, M. (2024). Responsible AI-based business process management and improvement. *Digital Society*, *3*(2), 23.
15. Priyadarshi, M. (2025). Autonomous AI agents transforming enterprise operations: From static automation to intelligent decision-making systems. *Journal Of Multidisciplinary*, *5*(7), 863-870.
16. Rosemann, M., Brocke, J. V., Van Looy, A., & Santoro, F. (2024). Business process management in the age of AI–three essential drifts. *Information Systems and e-Business Management*, *22*(3), 415-429.
17. Shinn, N., Cassano, F., Gopinath, A., Narasimhan, K., & Yao, S. (2023). Reflexion: Language agents with verbal reinforcement learning. *Advances in neural information processing systems*, *36*, 8634-8652.
18. Wang, L., Ma, C., Feng, X., Zhang, Z., Yang, H., Zhang, J., ... & Wen, J. (2024). A survey on large language model based autonomous agents. *Frontiers of computer science*, *18*(6), 186345.
19. Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., ... & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, *35*, 24824-24837.
20. Xi, Z., Chen, W., Guo, X., He, W., Ding, Y., Hong, B., ... & Gui, T. (2025). The rise and potential of large language model based agents: A survey. *Science China information sciences*, *68*(2), 121101.
21. Xie, T., Zhang, D., Chen, J., Li, X., Zhao, S., Cao, R., ... & Yu, T. (2024). Osworld: Benchmarking multimodal agents for open-ended tasks in real computer environments. *Advances in Neural Information Processing Systems*, *37*, 52040-52094.
22. Yang, J., Jimenez, C., Wettig, A., Lieret, K., Yao, S., Narasimhan, K., & Press, O. (2024). Swe-agent: Agent-computer interfaces enable automated software engineering. *Advances in Neural Information Processing Systems*, *37*, 50528-50652.
23. Yao, S., Chen, H., Yang, J., & Narasimhan, K. (2022). Webshop: Towards scalable real-world web interaction with grounded language agents. *Advances in Neural Information Processing Systems*, *35*, 20744-20757.
24. Yao, S., Yu, D., Zhao, J., Shafran, I., Griffiths, T., Cao, Y., & Narasimhan, K. (2023). Tree of thoughts: Deliberate problem solving with large language models. *Advances in neural information processing systems*, *36*, 11809-11822.
25. Zhou, S., Xu, F. F., Zhu, H., Zhou, X., Lo, R., Sridhar, A., ... & Neubig, G. (2024, May). Webarena: A realistic web environment for building autonomous agents. In *International Conference on Learning Representations* (Vol. 2024, pp. 15585-15606).