

Article

AI Agents and No-Code Tools in Accounting: A Case Study

Miguel Resende 

Governança, Competitividade e Políticas Públicas—Research Unit on Governance, Competitiveness and Public Policies, Higher Institute of Accounting and Administration, University of Aveiro, 3810-193 Aveiro, Portugal; miguelresende@ua.pt

Abstract

Advances in Artificial Intelligence (AI) and Large Language Models (LLMs) have transformed accounting by automating repetitive tasks and enhancing the efficiency of financial reporting. However, their implementation raises challenges related to bias, reliability, and professional adaptation. This article evaluates the comparative performance of three approaches to the vertical analysis of income statements: the traditional manual process, a specialized GPT model, and an AI agent integrating GPT with no-code automation tools. Using the Design Science Research (DSR) methodology, 150 experimental analyses were conducted to measure the execution time, variability, and process scalability. The results indicate that GPT substantially reduced execution time compared to the manual baseline, but still required significant human intervention. The AI agent achieved the greatest gains, reducing the average execution time by nearly 75%, while also demonstrating more stable performance and minimizing the repetitive workload. These findings provide empirical evidence that agent-based automation enhances both efficiency and reliability in accounting workflows, reinforcing its potential to reshape professional practice by reallocating human effort to validation and analytical tasks.

Keywords: Artificial Intelligence; Large Language Models; AI agents; accounting automation; financial reporting; productivity

JEL Classification: M41; M15; 033; C33



Academic Editor: David (Yulong) Liu

Received: 10 October 2025

Revised: 16 November 2025

Accepted: 21 November 2025

Published: 23 November 2025

Citation: Resende, M. AI Agents and No-Code Tools in Accounting: A Case Study. *FinTech* **2025**, *4*, 65. <https://doi.org/10.3390/fintech4040065>

Copyright: © 2025 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Technological advancements have significantly transformed the accounting paradigm, particularly with the advent of Artificial Intelligence (AI) and Large Language Models (LLMs). These technologies facilitate the automation of routine tasks and redefine the role of accountants, thus necessitating the acquisition of new analytical and strategic competencies [1]. The integration of LLMs into accounting has markedly enhanced operational efficiency, reshaping professionals' interpretation of financial information and requiring greater adaptation to emerging technological demands [2]. LLMs can automate processes such as bank reconciliation, data analysis, forecasting, and the preparation of corresponding reports [3]. However, their implementation poses certain challenges, including algorithmic bias (commonly referred to as hallucinations) and resistance from professionals [4]. While financial analysts and auditors tend to swiftly adopt these tools, accountants and tax consultants exhibit greater resistance owing to a lack of training and concerns regarding job displacement [5].

The increasing application of AI in accounting has created ethical and operational challenges. Some studies emphasize the responsible use of LLMs, aiming to mitigate risks such as model hallucinations and lack of transparency in processes and data usage [4]. Continuous professional training for accountants is essential to ensure oversight and regulatory compliance [5]. Digital transformation driven by LLMs represents not only technological advancement but also a paradigm shift in the accountant's role, making it more focused on data analysis and strategic consulting [6].

Building on these advances, recent studies have also explored the use of chatbots and AI agents in accounting, highlighting their potential to automate repetitive tasks and support financial reporting [7–12]. Although these technologies have demonstrated significant gains in efficiency and user acceptance, challenges such as data security, professional resistance, and the need for continuous training remain [6,12].

Against this backdrop, AI agents have emerged as a promising development. By integrating LLMs with no-code automation platforms, they enable end-to-end workflows that minimize human intervention and significantly reduce execution time. The objective of this article is to compare the efficiency and reliability of three alternative approaches to the vertical analysis of income statements, manual execution, specialized GPT, and AI agent automation. Using the Design Science Research (DSR) methodology and 150 experimental cases, we measured execution time, process stability, and scalability across the three approaches. The results demonstrate that the AI agent achieved the greatest efficiency gains, reducing the average execution time by nearly 75% relative to the manual baseline, while also improving consistency and process scalability.

This article contributes to the literature in several ways. First, it empirically validates the efficiency gains from integrating AI agents into accounting workflows, an area still underexplored in prior research. Second, it highlights the organizational implications of scalability, showing how automation can fundamentally reshape the economics of financial reporting services. Third, it underscores the importance of professional adaptation, as accountants must shift from manual execution to validation, interpretation, and advisory tasks. By demonstrating the feasibility and advantages of agent-based automation, the findings reinforce the strategic value of AI in enhancing both the productivity and professional role of accountants. Although recent research has explored the impact of LLMs and chatbots on accounting, empirical comparisons between manual, GPT-assisted, and agent-based automation remain scarce. Most prior studies are conceptual or limited to isolated tools, leaving a gap in the understanding of how fully automated agents perform in realistic accounting workflows.

This article is organized into five sections. Following this introduction, the second chapter presents a literature review. The third and fourth chapters detail the methodology and results, respectively, along with their discussion. The fifth chapter presents the main conclusions, limitations, and suggestions for future research.

2. Literature Review and Hypothesis Development

The adoption of LLMs in accounting redefines accountants' roles by automating repetitive tasks and demanding greater analytical and strategic capabilities [1]. Artificial Intelligence not only enhances operational efficiency but also transforms the interpretation and use of financial information, requiring advanced technological skills and critical thinking [2]. LLMs have demonstrated significant potential in bank reconciliations, financial reporting, and strategic and predictive analyses [3]. However, their implementation faces challenges, such as lack of regulation, algorithmic bias, and resistance from professionals [4]. Beyond automation, LLMs assist in interpreting accounting standards and formulating strategic insights, acting as support for professional judgment rather than replacing it [13].

Nevertheless, a critical analysis of their outputs is essential because of the risk of algorithmic bias, which may compromise the reliability of collected data, analyses, and subsequent decision-making [14]. During auditing, LLMs enhance the continuous analysis of financial data, facilitating fraud detection and anomaly identification [15]. However, the opacity of their decision-making processes raises concerns, highlighting the importance of frameworks that incorporate human validation and security layers [2,4,16]. Continuous training of accountants is crucial to ensure that they understand the potential and limitations of this technology [4,16]. Based on this discussion, the first research hypothesis is proposed:

H1: *GPT provides greater productivity than the manual process.*

The application of LLMs in accounting has been analyzed with a focus on prompt engineering techniques aimed at optimizing the performance of these models in complex tasks. Three primary approaches stand out, one-shot prompting, Chain of Thought (CoT), and Tree of Thoughts (ToT), each of which has different implications for the effectiveness of LLMs. One-shot prompting involves presenting a single example before requesting a response. Ref. [17] demonstrated that LLMs, such as GPT-3, can learn new tasks without requiring parameter reconfiguration. In accounting, this technique is used to automate classification processes, allowing a model to record transactions based on a single example. Ref. [18] further highlighted its application in financial statement analysis, reducing manual review time and delivering faster results. CoT prompting is more effective for tasks that require structured reasoning. Ref. [19] demonstrated that this technique breaks down problems into smaller steps, thereby improving the accuracy of the results. CoT prompting guides the model through sequential reasoning steps before presenting the final output. Refs. [20,21] emphasize that CoT is more appropriate for structured problems, such as financial statement analysis, where each item must be interpreted in relation to a total figure.

ToT prompting extends CoT by structuring reasoning into a tree format, thus enabling the exploration of multiple hypotheses before reaching a final decision. Ref. [22] found that this technique is particularly useful in forensic accounting and financial auditing, where evaluating diverse scenarios enhances the detection of irregularities. Ref. [23] indicated that ToT improves regulatory compliance accuracy by allowing models to adjust their responses based on new information.

Thus, the adoption of chatbots in accounting has been growing, driven by advancements in AI and LLMs. Ref. [24] emphasized that prompt engineering influences the AI model's response accuracy, impacting risk management and operational efficiency. Although chatbots have reduced the time required for financial operations, challenges, such as algorithmic bias, remain. Recent developments in the financial services sector show how AI agents reshape customer interaction, risk management, and operational efficiency [25]. These dynamics are expected to influence accounting, whereby AI-driven agents can automate compliance monitoring, reporting, and decision support. From this perspective, the following hypothesis was formulated:

H2: *The agent provides greater productivity than the manual process.*

Ref. [7] analyzed the application of AI in automated financial statement analysis and concluded that model customization and human supervision are essential for ensuring reliable analyses. The interpretation of highly specific accounting data remains a challenge requiring continuous chatbot adjustments to prevent errors and inaccurate conclusions. Ref. [8] specifically examines the impact of chatbots on accounting, highlighting their usefulness in automating repetitive tasks, such as bank reconciliation and invoice process-

ing. However, concerns about data security and professional resistance have hindered their widespread adoption. Ref. [9] demonstrated that chatbots can automate tasks such as tax calculations, bank reconciliations, and financial report generation, facilitating tax compliance and improving financial management. Ref. [10] studied a chatbot application for bank reconciliations and accounts' payable/receivable management, revealing a positive user perception, with users finding the tool intuitive and effective in reducing their manual effort. The authors estimated that automation could save up to 1500 working hours per year. Ref. [6] highlighted that chatbots can revolutionize accounts' receivable and payable reconciliation and financial reporting, allowing accountants to assume more strategic roles. However, professional resistance, data privacy, and regulatory constraints remain obstacles to their adoption. Ref. [11] analyzed the use of ChatGPT plus, 4.0 model in forensic accounting and demonstrated that AI can synthesize large volumes of information and detect suspicious financial patterns. However, limitations, such as errors in financial calculations and difficulties in interpreting accounting standards, reinforce the need for human oversight to prevent decision-making failures. Ref. [12] warns of the challenges faced by small and medium-sized enterprises in adopting new technologies, emphasizing the role of accountants in this transition and the necessity of continuous training for these professionals. Considering the relative performances of different AI-based approaches, a third hypothesis is proposed:

H3: *The agent provides greater productivity than GPT.*

3. Methodology

This article adopted the DSR methodology to provide a methodological bridge between problem-oriented design and empirical validation. Originally developed in information systems and software engineering, DSR emphasizes the creation and iterative evaluation of technological artifacts to address practical problems. Its relevance to accounting research lies in its capacity to translate technical innovations, such as AI and automation, into applied, measurable improvements in professional contexts. Following [26], this study applies DSR through iterative cycles of design, implementation, and evaluation to ensure both technical rigor and professional relevance. The purpose of employing DSR is twofold: (i) to design and implement technological artifacts that automate the vertical analysis of income statements (see Figure 1) and (ii) to empirically evaluate their impact on accountants' productivity and accuracy. The methodological rigor of DSR ensures that the proposed solutions are not only technically functional but also practically relevant to the accounting profession.

Three alternative methods for conducting vertical analysis were examined:

- A manual method is used to represent the traditional approach;
- Specialized GPT is a tailored Large Language Model developed through prompt engineering to support accountants in the preparation of executive summaries;
- An AI agent, which is a no-code workflow solution integrating GPT with automation platforms, was designed to minimize human intervention in repetitive tasks.

As of December 31, 2024 (in euros)

Income and Expenses	2024	2023
Sales and services rendered	€1,667,563.42	€1,883,449.65
Cost of goods sold and materials consumed	(€1,133,197.84)	(€1,301,561.37)
Supplies and external services	(€194,813.35)	(€172,647.82)
Personnel expenses	(€190,758.20)	(€188,908.78)
Impairment of receivables (losses / reversals)	(€24,780.68)	(€76,843.41)
Other income	€17,866.29	€16,704.00
Other expenses	(€20,254.70)	(€4,180.66)
Earnings before depreciation, financing costs, and taxes	€121,624.94	€156,011.61
Depreciation and amortization (expenses / reversals)	(€21,760.15)	(€29,082.25)
Operating result (before financing costs and taxes)	€99,864.79	€126,929.36
Interest and similar income	–	€9.50
Interest and similar expenses	(€44,490.35)	(€55,681.79)
Result before taxes	€55,374.44	€71,257.07
Income tax for the period	(€13,289.87)	(€16,802.93)
Net income for the period	€42,084.57	€54,454.14

Company statement – Trade of goods – IT products
Tax ID: 999999999

Figure 1. Income statement structure used as input.

3.1. Manual Method

The baseline process relies entirely on human effort. Accountants manually downloaded income statements in XLSX format, performed calculations in spreadsheets, drafted the textual analysis, and prepared the executive summary for clients. For the present article, this process was replicated 50 times to allow comparison with the other approaches, controlled by the Customer Relationship Management (CRM) system.

3.2. Specialized GPT Method

The second method employed a customized GPT model developed using the OpenAI environment. Accountants uploaded income statements in PDF format, which the model processed to generate a structured executive summary. The accountant's role was limited to validating and, if necessary, correcting the draft before transmitting it to the client.

The GPT artifact was engineered using a structured prompt design based on CoT logic encoded in XML blocks, enabling step-by-step reasoning and consistent narrative generation. Prior to the experiment, the model was pre-tested to verify the accuracy, reproducibility, and stability of the outputs. Its credibility was established by benchmarking against manually computed vertical analyses and executive summaries prepared by accounting professionals.

Although this method reduces the calculation and drafting time, it still requires significant manual handling of inputs and outputs.

3.3. AI Agent Method

The third approach employs an AI agent built on the MAKE platform, which enables the orchestration of multiple Google services in combination with OpenAI and Google API. This configuration allowed for end-to-end automation of the vertical analysis process, significantly reducing manual intervention.

The AI agent reproduced the same reasoning framework as the GPT artifact through embedded XML-structured CoT prompts, while integrating automated validation loops and error-handling routines. Before deployment, the workflow was tested in a controlled

environment to ensure that the automated operations produced outputs consistent with GPT and manual benchmarks. This preliminary validation confirmed functional reliability and reduced the likelihood of execution errors during the experimental phase.

The workflow was divided into two automated flows:

Flow 1—Analysis and Archiving (Figure 2)

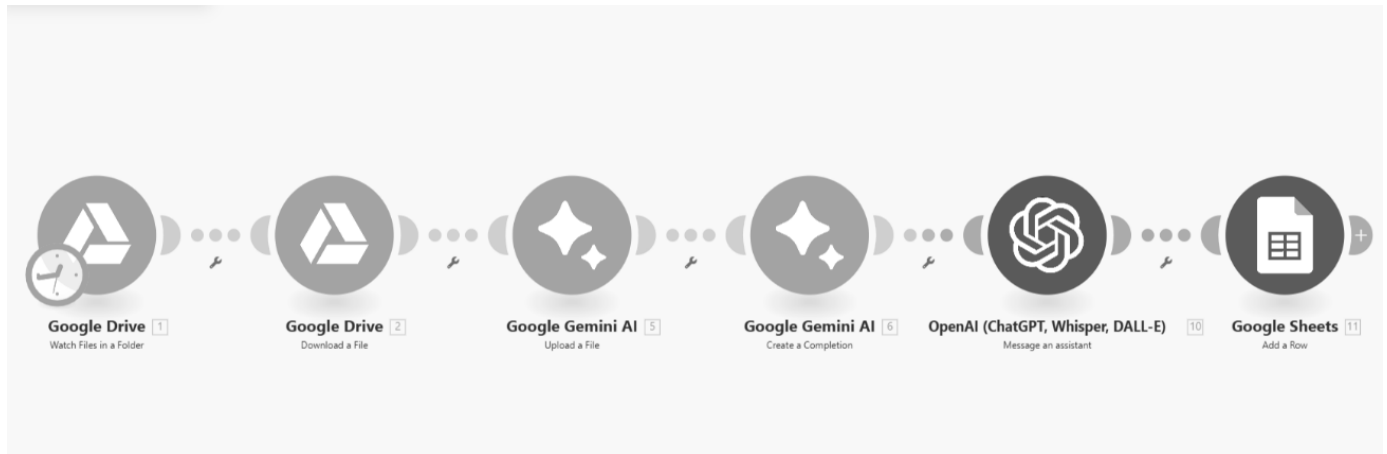


Figure 2. Automated workflow for analysis and archiving implemented in MAKE. The process triggers AI-based vertical analysis and saves outputs to Google Sheets.

1. Income statements in PDF format were uploaded to a designated Google Drive folder;
2. The AI agent automatically detected each new file, processed it sequentially, and performed a vertical analysis using structured prompts via OpenAI and Gemini's model;
3. The resulting executive summaries were generated and stored in Google Sheets for verification;
4. At this stage, the accountant reviewed and, if necessary, corrected the summaries to ensure accuracy and compliance with the professional standards.

Flow 2—Client Communication (Figure 3)

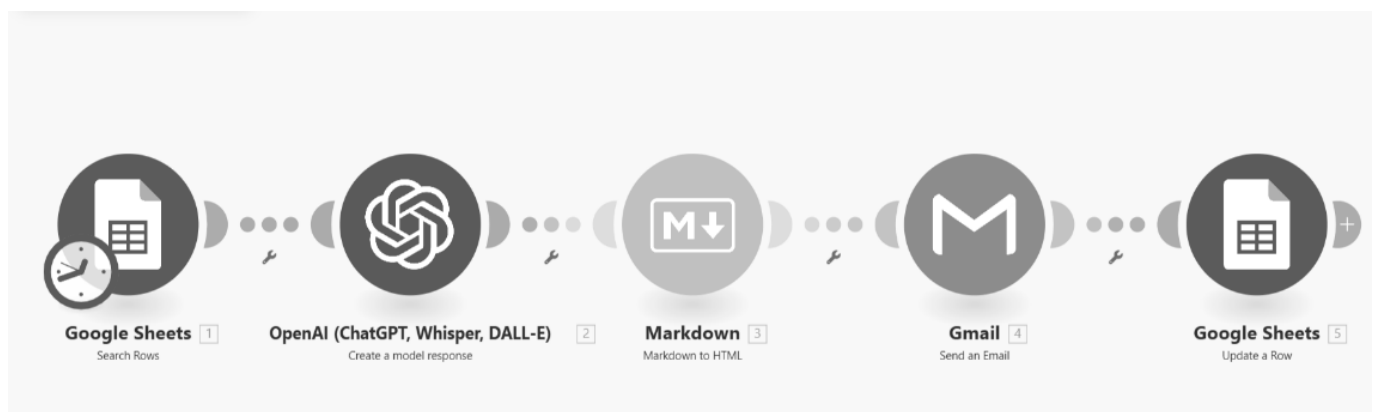


Figure 3. Automated client communication workflow. Once validated, summaries are formatted and sent automatically via Gmail, with logs recorded in Google Sheets.

1. Once validated, the summaries were formatted in Markdown and converted to client-ready reports;
2. The system automatically triggered Gmail to dispatch the reports directly to the respective clients;
3. Simultaneously, Google Sheets was updated to register the completion status of each case.

This dual-flow design not only automated the analytical and reporting tasks but also introduced a high degree of scalability, as multiple files could be processed in batches without additional human effort. To ensure comparability, the same prompt engineering strategy applied in the specialized GPT method was reused within the MAKE workflows. This allowed for direct benchmarking of outputs and processing times between the GPT-only and AI agent configurations.

By integrating data storage, analysis, validation, formatting, and communication into a seamless pipeline, the AI agent method minimized repetitive tasks and reduced cognitive fatigue. Thus, the accountant's role was concentrated on higher-value activities, validation, and professional judgment, whereas the system handled operational and logistical processes.

3.4. Application of the DSR Framework

The DSR methodology was applied in five iterative phases, adapted to the accounting context to ensure both technical rigor and professional relevance. The experimental cases are drawn from 50 Portuguese firms operating across a variety of sectors. These include professional and technical services (accounting, design, event organization, and topography), financial and insurance activities (insurance mediation), industrial and maintenance services, hospitality and food services (restaurants and coffee bar), and commercial and construction activities, encompassing retail and wholesale trade, real estate transactions, commission-based trade, wine and domestic goods sales, and building materials commerce. Based on the NACE Rev. 2 classification, approximately 64% of the firms belong to service-related activities—categories M (Professional, Scientific and Technical Activities), N (Administrative and Support Services), K (Financial and Insurance Activities), Q (Health Services), and I (Accommodation and Food Services)—while 36% operate in commerce and construction, corresponding to categories G (Wholesale and Retail Trade), F (Construction), and L (Real Estate).

All firms qualify as small or medium-sized enterprises (SMEs) under the European definition, with fewer than 50 employees and either less than EUR 10 million in annual turnover or under EUR 2 million in total assets. Each company applies the Portuguese Accounting Standardization System (SNC), which is IFRS-based, to ensure consistency in accounting treatment across cases. Although statistical representativeness was not the primary goal of this study, the sample reflects a realistic distribution of small and medium-sized enterprises across service and commercial sectors. Given their operational diversity and shared IFRS-based reporting framework, the sample can be considered methodologically representative for evaluating and comparing process efficiency among the three tested approaches. This homogeneous regulatory context, combined with organizational diversity in size and sector, allows for the robust testing of workflow automation while maintaining the external validity of the findings:

1. **Problem Identification:** Analysis of CRM logs revealed that a manual vertical analysis of income statements required approximately 10 min per report. This highlights a clear opportunity for automation that aims to reduce repetitive effort and allow accountants to focus on value-added tasks.
2. **Design of the Solution:** Two artifacts were developed: (i) a specialized GPT model, configured through structured prompt engineering, and (ii) an AI agent, implemented through no-code automation (MAKE) integrated with the OpenAI API. Both artifacts were designed as follows:

- a. Extract relevant parameters from standardized income statement.
 - b. Execute vertical analysis by calculating each item as a percentage of total revenue.
 - c. Produce an executive summary structured into three parts: overall vertical analysis, detailed item-level analysis, and an integrated summary with recommendations. The CoT prompting strategy was adopted to ensure step-by-step reasoning, consistent with accounting logic and numerical interpretation.
3. Validation of the Design: Validation focused on ensuring both computational accuracy and interpretive consistency. Outputs from the GPT and AI agent were benchmarked against manually calculated cases to verify adherence to accounting principles and the reliability of the generated executive summaries.
4. Implementation: Specialized GPT was deployed through XML-based structured prompts with embedded Python code execution capabilities (OpenAI Code Interpreter environment, Python 3), whereas the AI agent leveraged MAKE workflows to orchestrate Google Drive, Google Sheets, Gmail, Gemini, and OpenAI services. This dual implementation enabled a systematic comparison between semi-automated (GPT) and fully automated (AI agent) workflows.
5. Evaluation: To ensure robustness, each of the three methods (manual, GPT, and AI agent) was executed 50 times across different companies over three months, resulting in 150 experimental cases. The execution time, stability, and accuracy of the outputs were systematically measured. Human validation remains mandatory in all workflows to safeguard their professional reliability. This phase also served as a practical test of implementation, capturing not only statistical differences but also qualitative impacts on accountants' workload and scalability.

4. Results and Discussion

The execution time and stability data were automatically recorded using the CRM system logs used during the experimental runs described in Section 3.4. These logs capture the start and end timestamps for each accounting report. The exported datasets were processed using JAMOV (2.6.44) statistical software to compute descriptive statistics and perform comparative analyses across the three approaches.

4.1. Graphical Analysis

Figure 4 shows that the graphical analysis highlights the distinct performance profiles across the three approaches. The manual method exhibited a gradual upward drift in the execution time, reflecting the cumulative burden of repetitive tasks and the influence of human fatigue. Specialized GPT reduced execution times considerably, but retained a mild drift, as accountants were still required to upload, copy, and dispatch reports manually. By contrast, the AI agent delivered the lowest and most stable execution times, showing no significant signs of drift or fatigue. This stability reflects a higher level of process automation and a reduction in repetitive human tasks, enabling accountants to concentrate primarily on validation and professional judgment.

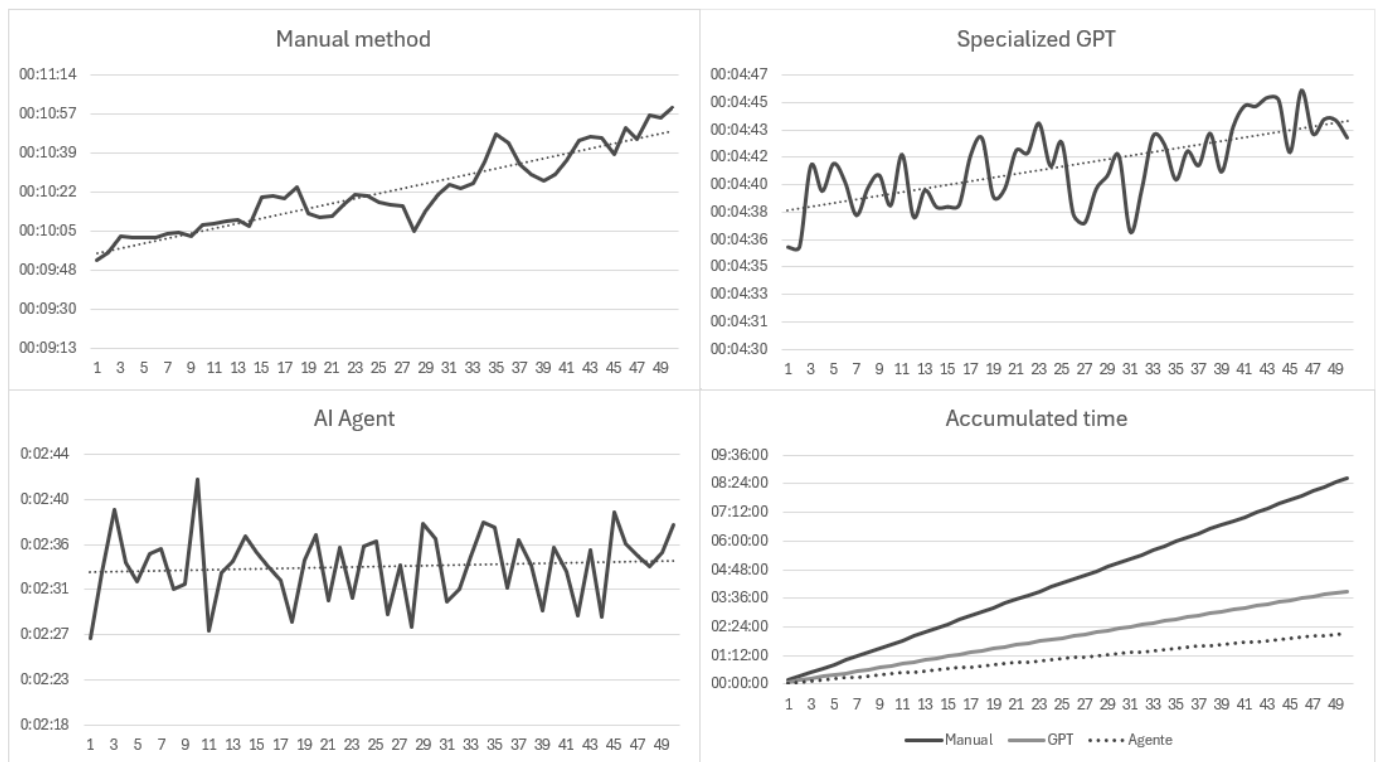


Figure 4. Graphical analysis highlights distinct performance profiles across the three approaches. Source: experiment log data extracted from the CRM system, processed and created in Microsoft Excel.

4.2. Qualitative Assessment of Methods

Beyond the quantitative gains, qualitative insights further differentiate the approaches.

4.2.1. Specialized GPT Method

- Strengths: Reduced drafting time and summaries passed verification.
- Limitations: A high degree of manual work remained (uploading, copying, and sending each report individually), and full human presence was required throughout the process.

4.2.2. AI Agent Method

- Strengths: Substantial reduction in manual workload; reports uploaded collectively to a single folder; lower fatigue; greater focus on analysis and validation; and elimination of human intervention in two of the most repetitive and time-consuming phases.
- Limitations: Manual uploading of financial statements into Google Drive folder is still required. Future integration with the TOOnline ERP platform (cloud-based SaaS, versioning not applicable) could eliminate this residual step.

4.3. Descriptive Statistics

Table 1 reports the descriptive statistics of execution times across the three process configurations. The manual workflow required, on average, 622.4 s (≈ 10 min 22 s) per analysis, whereas the specialized GPT reduced this to 281.2 s (≈ 4 min 41 s). The AI agent achieved the lowest mean execution time of 153.5 s (≈ 2 min 33 s). The median values closely matched the mean values in all three cases, confirming the absence of strong asymmetries in the distributions.

Table 1. Description of the sample.

Methods	N	Min.	Max.	Average	Median	Standard Deviation
Manual method	50	00:09:52	00:11:00	00:10:22	00:10:20	00:00:17
Specialized GPT	50	00:04:36	00:04:46	00:04:41	00:04:41	00:00:02
AI agent	50	00:02:27	00:02:42	00:02:33	00:02:34	00:00:03

Time values format (hh:mm:ss).

Regarding variability, the manual process had the highest standard deviation ($SD = 17.4$ s), indicating greater inconsistency across repetitions. The GPT exhibited markedly greater stability ($SD = 2.4$ s), whereas the AI agent achieved both low mean times and consistently stable performance ($SD = 3.3$ s). These results suggest that automation accelerates execution and improves the reliability of repetitive accounting tasks.

The observed ranges reinforce this pattern. The manual method varied between 00:09:52 and 00:11:00, whereas GPT narrowed this range considerably (00:04:36 to 00:04:46). The AI agent further compressed it (00:02:27 to 00:02:42). This tightening of the range underscores the superior predictability of automated methods compared with manual procedures.

Pairwise comparisons highlight the scale of these improvements. Transitioning from the manual to the GPT workflow yielded a mean saving of 341.3 s (≈ 5 min 41 s), representing a 55% reduction. Moving from GPT to the AI agent provided an additional gain of 127.7 s (≈ 2 min 08 s), which was equivalent to a further 45% reduction. In total, the AI agent reduced the execution times by 468.9 s (≈ 7 min 49 s), corresponding to a 75% improvement compared to the manual baseline.

Finally, the normality of the distributions was assessed using the Shapiro–Wilk test. The results were as follows: manual ($W = 0.954$, $p = 0.051$), GPT ($W = 0.984$, $p = 0.722$), and agent ($W = 0.982$, $p = 0.644$). In all cases, the p -values were above the threshold of 0.05, indicating that none of the datasets showed significant deviations from normality. Therefore, the assumption of normality was accepted, supporting the application of parametric tests, such as the paired-samples t -test, in subsequent analyses.

4.4. Analysis and Discussion of Results

Table 2 presents the results of parametric t -tests for the three hypotheses. In all cases, the results are highly significant ($p < 0.001$), providing strong statistical evidence to support the proposed hypotheses.

Table 2. Hypothesis testing results.

Hypothesis	Parametric t-Statistic	GI	p -Value
H1. GPT provides greater productivity than the manual process.	153	49	<0.001
H2. The agent provides greater productivity than the manual process.	223	49	<0.001
H3. The agent provides greater productivity than GPT.	192	49	<0.001

For H1, comparing the manual process with the specialized GPT, the test produced a t -statistic of 153 ($df = 49$, $p < 0.001$). This confirms that GPT substantially increases productivity relative to manual execution. The effect size is extremely large, indicating that the observed improvement is not only statistically significant but also of practical importance for accounting operations.

For H2, the comparison between the manual method and the AI agent yielded the largest difference, with a t -statistic of 223 ($df = 49$, $p < 0.001$). This demonstrates that the

AI agent delivers the most pronounced efficiency gains, reducing both execution time and human workload. This result underscores the transformative potential of full process automation for routine accounting tasks.

For H3, the head-to-head comparison between the GPT and AI agent resulted in a t-statistic of 192 ($df = 49$, $p < 0.001$). This strong significance confirms that the AI agent outperforms the GPT, validating the expectation that integrated automation (AI agent) provides greater productivity than a semi-automated, GPT-only workflow.

Taken together, these results provide robust support for all the three research hypotheses (H1–H3). Importantly, the consistently high t-values indicate that the differences between the methods are not marginal but rather substantial.

The shift toward AI-augmented workflows implies that accountants increasingly require hybrid competencies, combining analytical reasoning, data interpretation, and AI oversight. Continuous professional education and curricular integration of automation tools are essential to ensure that practitioners can critically evaluate outputs and uphold ethical and regulatory standards.

4.5. Ethical and Regulatory Considerations

The implementation of AI-driven workflows in accounting requires strict consideration of data privacy, model transparency, and professional ethics. All experimental cases were conducted using real but anonymized financial data, in which the names of companies and any identifiable details were replaced with internal numeric identifiers to ensure confidentiality. Data processing followed the organization's internal privacy policy, aligned with the principles of the EU General Data Protection Regulation (GDPR).

The AI tools employed in this study, OpenAI GPT and Google (via API), operate under certified information-security frameworks. OpenAI's API environment is covered by SOC 2 Type 2 reporting and supports customer compliance with the GDPR and CCPA, including regular third-party penetration testing and a formal Data Processing Addendum for enterprise clients. Similarly, for Google, Cloud services comply with multiple international standards (ISO 27001, 27017, 27018, 27701; ISO 42001; SOC 1–3; HIPAA; CSA STAR; BSI C5) [27], ensuring data protection and operational integrity.

From an ethical standpoint, this article acknowledges that algorithmic decision systems must be transparent and auditable. All AI-generated outputs were subject to human validation before being accepted, thereby ensuring professional accountability and mitigating the risks of bias, hallucinations, or misinterpretation.

5. Conclusions

This article examines the comparative performance of three approaches to the vertical analysis of financial statements: the traditional manual method, a specialized GPT model, and a fully automated AI agent. The objective is to assess whether automation, particularly through AI agents, can significantly improve efficiency, reduce human workload, and enhance scalability in routine accounting tasks.

The results demonstrated substantial differences between the methods. The manual process, which required approximately 10 min per analysis, reflected the cumulative inefficiencies of the repetitive tasks. The introduction of a specialized GPT model reduced this time to 5 min 41 s, yielding a 55% efficiency gain. The AI agent further reduced average execution time to 2 min 08 s, a 45% reduction in relation to the GPT, and achieved an overall improvement of approximately 75% relative to the manual baseline. Statistical tests confirmed that these differences were highly significant, which was supported by the three research hypotheses. Importantly, the AI agent not only accelerated processing, but

also delivered greater stability, eliminating the drift and variability observed in manual and GPT-supported workflows.

From a professional perspective, this article contributes to the accounting literature by illustrating the tangible productivity benefits of integrating AI agents into the financial reporting processes. Beyond efficiency gains, a reduction in manual interventions decreases cognitive load and fatigue, allowing accountants to focus on validation and higher-value analytical tasks. Furthermore, the scalability of the AI agent introduces transformative potential; batch processing enables firms to manage large client portfolios without a proportional increase in manual effort, fundamentally reshaping the economics of accounting services.

This article has some limitations. The experimental design was conducted in a controlled environment using 50 Portuguese firms operating in the service and retail sectors, which may limit the generalizability of the findings to other contexts, sectors, or regulatory regimes. However, this controlled setup was necessary to ensure that all 150 analyses were performed under identical accounting and reporting conditions, allowing a valid comparison of the three methodological approaches (manual, GPT, and AI agent) without introducing variability from heterogeneous data structures or firm characteristics. Out-of-sample validation was not performed, and the analysis focused primarily on quantitative efficiency metrics, execution time, stability and scalability, rather than on broader organizational or behavioral outcomes. Human validation was still necessary to ensure compliance and contextual accuracy. The workflow was only tested for structured and repetitive tasks. Additionally, current workflows still require manual uploading of financial statements, indicating the need for future integration between AI agents and ERP or other management platforms.

Future research should extend this line of inquiry by exploring the applicability of agent-based automation to complex, judgment-driven areas of accounting, such as audit analytics, fraud detection, and strategic advisory functions. It should also incorporate out-of-sample validation to evaluate whether the observed efficiency and stability persist when applied to unseen datasets or alternative financial statement formats across different firms and sectors. Investigating the integration of AI agents directly into ERP systems through APIs may further reduce residual manual steps, pushing automation closer to real-time, end-to-end accounting solutions.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The anonymized datasets generated and analyzed during the study are available from the corresponding author upon reasonable request.

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Kroon, N.; Alves, M.C.; Martins, I. The Impacts of Emerging Technologies on Accountants' Role and Skills: Connecting to Open Innovation—A Systematic Literature Review. *J. Open Innov. Technol. Mark. Complex.* **2021**, *7*, 163. [\[CrossRef\]](#)
2. Anica-Popa, I.-F.; Vrîncianu, M.; Anica-Popa, L.-E.; Cismasu, I.-D.; Tudor, C.-G. Framework for Integrating Generative AI in Developing Competencies for Accounting and Audit Professionals. *Electronics* **2024**, *13*, 2621. [\[CrossRef\]](#)
3. Li, H.; Vasarhelyi, A. Applying Large Language Models in Accounting: A Comparative Analysis of Different Methodologies and Off-the-Shelf Examples. *J. Emerg. Technol. Account.* **2024**, *21*, 133–152. [\[CrossRef\]](#)
4. Mwachikoka, T. Effects of Artificial Intelligence on Financial Reporting Accuracy. *World J. Adv. Res. Rev.* **2024**, *23*, 1751–1767. [\[CrossRef\]](#)

5. Toumeh, A.A. Assessing the Potential Integration of Large Language Models in Accounting Practices: Evidence from an Emerging Economy. *Future Bus. J.* **2024**, *10*, 82. [\[CrossRef\]](#)
6. Odonkor, B.; Kaggwa, S.; Uwaoma, P.U.; Hassan, A.O.; Farayola, O.A. The Impact of AI on Accounting Practices: A Review—Exploring How Artificial Intelligence Is Transforming Traditional Accounting Methods and Financial Reporting. *World J. Adv. Res. Rev.* **2024**, *21*, 172–188. [\[CrossRef\]](#)
7. Noels, S.; Merlevede, A.; Fecheyr, A.; Vanhalst, M.; Meerlaen, N.; Viaene, S.; De Bie, T. Automated Financial Analysis Using GPT-4. In *Lecture Notes in Artificial Intelligence, Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD 2023)*, Turin, Italy, 18–22 September 2023; De Francisci Morales, G., Perlich, C., Ruchansky, N., Kourtellis, N., Baralis, E., Bonchi, F., Eds.; Springer: Cham, Switzerland, 2023; Volume 14175, pp. 345–349. [\[CrossRef\]](#)
8. Resende, M.O. Impacto dos Chatbots na Contabilidade: Aplicações, Benefícios e Desafios. In Proceedings of the 10th GECAMB/P-CSEAR & XXIV Grudis Conference and Doctoral Colloquium, Bragança, Portugal, 23–25 January 2025.
9. Zadorozhnyi, Z.-M.; Muravskiy, V.; Muravskiy, V.; Shevchuk, A.; Pochynok, N.; Majda, M. Application of Chatbots with Artificial Intelligence in Accounting. In Proceedings of the 13th International Conference on Advanced Computer Information Technologies (ACIT), Wrocław, Poland, 21–23 September 2023; pp. 196–200. [\[CrossRef\]](#)
10. Bavaresco, R.S.; Nesi, L.C.; Barbosa, J.L.V.; Antunes, R.S.; Righi, R.R.; Costa, C.A.; Vanzin, M.; Dornelles, D.; Junior, S.C.; Gatti, C.; et al. Machine Learning-Based Automation of Accounting Services: An Exploratory Case Study. *Int. J. Account. Inf. Syst.* **2023**, *49*, 100618. [\[CrossRef\]](#)
11. Street, D.A.; Wilck, J. “Let’s Have a Chat”: Applying ChatGPT and Other Large Language Models to the Practice of Forensic Accounting. *SSRN Electron. J.* **2023**. Available online: <https://ssrn.com/abstract=4351817> (accessed on 30 April 2025).
12. Resende, M. Impacts of Accounting Digitization on the Profession of Accountants and Productivity: A Case Study. *Eur. J. Appl. Bus. Manag.* **2024**, *10*, 115–141. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Huang, X. Analysis of the Impact of Artificial Intelligence on Enterprise Financial Accounting Work. *Trans. Econ. Bus. Manag. Res.* **2024**, *7*, 44–48. [\[CrossRef\]](#)
14. Yoo, M. How Much Should We Trust Large Language Model-Based Measures for Accounting and Finance Research? *Univ. Penn. SSRN Electron. J.* **2024**. [\[CrossRef\]](#)
15. Antwi, B.; Adelakun, B.; Eziefule, A. Transforming Financial Reporting with AI: Enhancing Accuracy and Timeliness. *Int. J. Adv. Econ.* **2024**, *6*, 205–223. [\[CrossRef\]](#)
16. Barna, L.E.L.; Hurducaci, C.C. The Impact of Using Artificial Intelligence and ERP Systems in the Work of Accounting Professionals and Auditors. *Ann. Univ. Oradea Econ. Sci.* **2024**, *33*, 246–258. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Brown, T.B.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. Language Models Are Few-Shot Learners. *arXiv* **2020**, arXiv:2005.14165. [\[CrossRef\]](#)
18. Fung, S.; Wong, M.; Tan, C.W. Chain-of-Thoughts Prompting with Language Models for Accurate Math Problem-Solving. In Proceedings of the 2023 IEEE MIT Undergraduate Research Technology Conference (URTC), Cambridge, MA, USA, 6–8 October 2023; pp. 1–5. [\[CrossRef\]](#)
19. Wei, J.; Wang, X.; Schuurmans, D.; Bosma, M.; Ichter, B.; Xia, F.; Chi, E.H.; Le, Q.V.; Zhou, D. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. *arXiv* **2023**, arXiv:2201.11903. [\[CrossRef\]](#)
20. Duarte, R.D. Aprenda a Técnica Thought Chain Prompting na Análise de Balanço e DRE com ChatGPT. Available online: <https://www.robertodiasduarte.com.br/aprenda-a-tecnica-thought-chain-prompting-na-analise-de-balanco-e-dre-com-chatgpt/> (accessed on 30 April 2025).
21. Duarte, R.D. Guia para Aplicação das Técnicas Chain of Thought e Tree of Thoughts em Prompts. Available online: <https://www.robertodiasduarte.com.br/guia-para-aplicacao-das-tecnicas-chain-of-thought-e-tree-of-thoughts-em-prompts> (accessed on 30 April 2025).
22. Long, J. Large Language Model Guided Tree-of-Thought. *arXiv* **2023**, arXiv:2305.08291. [\[CrossRef\]](#)
23. Chen, B.; Zhang, Z.; Langrené, N.; Zhu, S. Unleashing the Potential of Prompt Engineering in Large Language Models: A Comprehensive Review. *arXiv* **2023**, arXiv:2310.14735. [\[CrossRef\]](#)
24. Dhar, A.; Datta, A.; Das, S. Analysis on Enhancing Financial Decision-Making Through Prompt Engineering. In Proceedings of the 7th International Conference on Electronics, Materials Engineering & Nano-Technology (IEMENTech), Kolkata, India, 18–20 December 2023; IEEE: New York, NY, USA, 2023. [\[CrossRef\]](#)
25. Fujitsu. AI Agents and the Transformation of the Financial Industry. Available online: <https://global.fujitsu/en-global/insight/tl-aiagents-financial-industry-20250418> (accessed on 30 April 2025).

26. Zaiden, F.; Bax, M.; Parreiras, F. Design Science Research: Application in a Research and Development Project. In Proceedings of the 13th International Conference on Information Systems & Technology Management (CONTECSI 2016), São Paulo, Brazil, 1–3 June 2016. [[CrossRef](#)]
27. ISO/IEC 27001:2022; Information Security, Cybersecurity and Privacy Protection — Information Security Management Systems — Requirements. International Organization for Standardization: Geneva, Switzerland, 2022.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.