NEW APPROACH FOR CLASSIFYING UNSTRUCTURED DATA TO UNDERSTAND THE INFLUENCE OF NEW TECHNOLOGIES IN FUTURE ERP DEVELOPMENT

Christian Geß*
ULSIT,
Bulgaria
c.gess@unibit.bg

Pepa Vl Petrova
ULSIT,
Bulgaria
p.petrova@unibit.bg

Georgi Dimitrov
ULSIT,
Bulgaria
g.dimitrov@unibit.bg

Abstract

The rapid development of emerging technologies is changing the business world every day. Process automation, artificial intelligence (AI) consultants, mobile devices, cloud technologies, no-code programming, and many more are influencing the requirements for business systems. The current study aims to compare human experts and generative AI (GenAI) in classifying unstructured text to identify the emerging technologies that will have the greatest impact on the future development of enterprise resource planning systems. The basis for the comparison is statistical indicators such as the average Jaccard index and Krippendorff's alpha, which validate the classification from both approaches.The results indicate that GenAI holds potential for process automation, whereas human classification is more closely aligned with statistical indicators. GenAI has difficulties with contextual nuances and specific categories. Hence, a four-step funnel-based framework for the efficient classification of unstructured text was developed,

*Corresponding Author

which integrates GenAI for initial structuring and training while emphasizing the indispensable role of human supervision for quality assurance to leverage automation while maintaining the validity of the results. This approach significantly reduces manual effort while maintaining reliable analysis.

Key Words

Enterprise resource planning (ERP) systems; text classification; generative artificial intelligence (GenAI); technological advancements; business administration

INTRODUCTION

Enterprise resource planning (ERP) systems have emerged as one of the most critical enterprise software platforms within the contemporary business landscape. Being aware of prevailing technological developments and keeping an eye on future trends is crucial, for example, when selecting a suitable ERP system or making investment decisions in innovative technologies. Therefore, it is essential to scientifically analyze the current literature, journal articles, and expert discussions to remain current.

Researchers who want to investigate which relevant technologies may influence the further development of ERP systems must first gather different types of source material and analyze it. These texts are typically unstructured, such as books, research papers, and expert interviews. The second step is to transform this unstructured data into a format that enables the extraction of valuable, actionable insights.

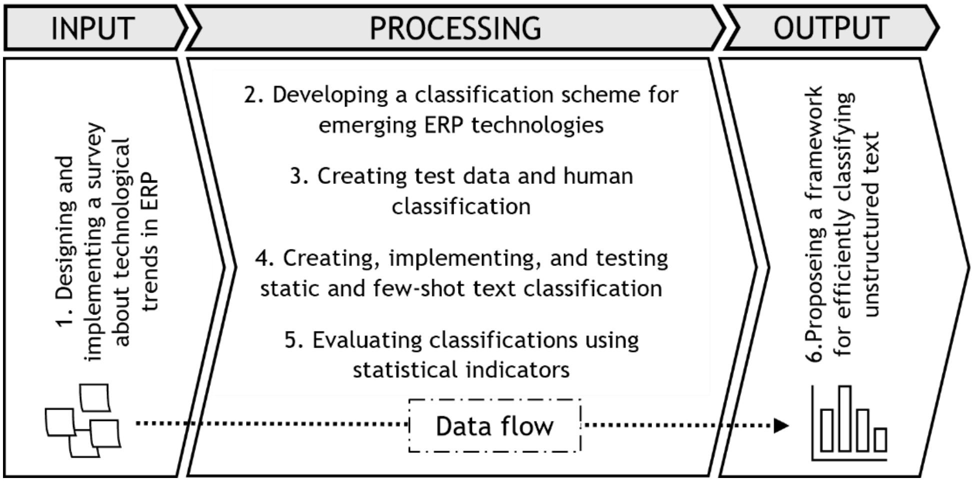
Unstructured data usually requires significant effort for researchers to perform data classification and evaluation, in contrast to structured data, which are often directly quantifiable. Reading texts carefully, performing comprehension, taking notes, and conceptually linking the sources are time-consuming tasks, especially when the sources are, for example, expert interviews, which integrate subjective opinions, assessments, future prognoses, and personal experiences. As a result, the context is often dynamic and interactive. Thus, classifying data can be exceptionally demanding, which often poses a problem for researchers, who must invest considerable time in classifying and analyzing it before they can obtain relevant findings.

Therefore, this article presents a new and up-to-date approach for minimizing the time spent classifying data without compromising quality. This novel approach aims to assist researchers in classifying unstructured texts by understanding the impact of modern technologies on future ERP development. The proposal integrates current technologies such as generative AI (GenAI) and provides recommendations for action, as well as concrete assistance for researchers.

METHODS AND RESEARCH APPROACH

This article follows the research design shown in Figure 1, as explained in the conceptual overview section below.

Figure 1: Research design



Conceptual overview

The sequence of six tasks, presented in Figure 1, forms the conceptual and logical framework of this article and is grouped into three stages: input, processing, and output. Firstly, the input stage encompasses the design and implementation of a survey about technological trends in ERP. Second, the processing stage comprises several important steps, which start with developing a classification scheme for emerging ERP technologies, followed by creating testing data and human classification. Third, the creation, implementation, and testing of static and few-shot text classification are conducted, which culminates in the evaluation of these classifications by applying statistical indicators. Fourth, the output stage delivers a comprehensive outcome built on the preceding stages and their relevant findings. A proposal for a novel methodology is subsequently presented. This methodology seeks to combine GenAI with traditional methods to achieve better and faster results for classifying non-structured text.

Designing and implementing a survey about technological trends in ERP

The survey aimed to confirm the importance of technologies expected to be significant in the ERP environment both now and in the future. These technologies were validated by a literature analysis encompassing 35 papers published between 2020 and 2024. These studies specifically examined the impact of at least three distinct technologies on ERP systems. Table 1 shows some of the survey participants' characteristics.

Table 1: Survey participants (*n* = 37)

ERP experience	
1–2	10
2–5	11
>5 years	16
Company size	
<50	8
50–100	2
100–500	9
> 500 employees	18

The participants' engagement with ERP software was also investigated beyond ensuring a balanced distribution while considering both experience and company size. Approximately one-third of the participants were traditional ERP system users. In addition, various other experts, including ERP system administrators, developers, and project managers, participated in the survey. The outcome was a well-balanced mix of ERP experience with varying company sizes and areas of responsibility. Consequently, gathering relevant domain knowledge and practical experience to develop a relevant and functional classification scheme in the ERP context was ensured through precise survey questions.

The survey was conducted online and designed to gather comprehensive expert opinions for the development of a classification scheme for emerging ERP technologies. Structured questions with Likert scales were used to identify trends regarding the influence of specific technologies on ERP.
In addition, binary questions (Yes/No) were used for specific decisions, supplemented by free-text questions to capture detailed opinions and additional perspectives.

The questions were grouped thematically, with both mandatory and optional questions to allow the participants to have flexibility in their responses while ensuring high data quality.

Several factors influenced the design and scope of the survey. Since the process of assessing free-text responses, understanding, and accurate labeling key technologies is inherently time-consuming, this constraint consequently necessitated a pragmatic reduction in the number of participants. Secondly, to ensure the representativeness of the findings, it was imperative to identify the appropriate target audience of experts. These individuals needed to possess the prerequisite knowledge to provide insightful answers regarding the future of ERP. Thirdly, the selection process itself demanded considerable effort to identify suitable individuals from each relevant group to participate in the questionnaire.

Overall, the chosen sample size was deemed sufficient to generate a qualitatively rich and diverse dataset despite these challenges. This dataset was intended for the exploratory development of a classification scheme while maintaining the practicality of both data collection and subsequent analysis.

Developing a classification scheme for emerging ERP technologies

The findings derived from the study served as the input for subsequent stages. Upon commencing the processing stage, the study results were thoroughly screened, and key technologies and keywords were extracted. Particular attention was paid during this process to ensuring the comprehensive integration of expert knowledge, with the aim of deriving precise category definitions and pertinent keywords. Subsequently, the findings pertaining to each identified technology were consolidated and clearly presented.

Figure 2: Classification scheme for emerging ERP technologies

The results of this work are presented in Figure 2, which provides a structured overview of a wide range of key technologies and concepts. In the context of this paper and due to its great importance in the ERP area, user experience is considered a technology that utilizes design principles, methods, and tools to create intuitive, efficient, and pleasant interfaces that increase acceptance and usability. These categories and technologies are shown, along with a brief description of the category. In addition, the gray boxes contain a list of keywords that indicate the technology. This classification scheme provided a guideline for recognizing the most important technologies in the domain of ERP systems. It facilitated the assignment of specific codes to text passages in the following sections of this research. The codes referred to the 20 relevant emerging technologies in the field of ERP systems.

Subsequently, a literature analysis was conducted to validate the relevance of the technologies outlined in the classification scheme for emerging ERP technologies. This analysis encompassed 35 papers, exclusively drawn from German and English literature published between 2020 and 2024, which specifically examined the impact of at least three distinct technologies on ERP systems.

Creating test data and human classification

A compilation of 100 examples from various sources was created to test the classification scheme with different approaches. Below are excerpts from texts dealing with technological trends in the field of ERP. The sources were written down sentence by sentence and numbered from 1 to 100, as shown in Table 2.

Table 2: Test dataset

No.	Sentence	Source
1	“Recent advancements in machine learning (ML), especially deep learning and ensemble techniques, have significantly improved the accuracy of predictive models.”	(Mhaskey, 2025)
...	…	
25	“The transformative impact of AI-powered invoice automation on financial operations demonstrates the significant potential of intelligent technology integration in modern business environments.”	(Onteddu, 2025)
…	…	
38	“The evidence available suggests that massive data and analytics are currently at the core of ERP systems and many related sectors.”	(Kaulwar, 2025)
…	…	
85	“Enterprise Resource Planning (ERP) systems are universally used to automate and manage business processes.”	(Sunmola and Lawrence 2024)
…	…	
100	“Previous research has been focused on critical success factors of ERP or critical success factors of blockchain.”	

Three text classifications were conducted within the scope of this article. The first was performed by a human evaluator who assigned the test data to categories sentence by sentence using the previously created classification scheme for emerging ERP technologies, as shown in Figure 2.

The human evaluator in such research must be an ERP expert with a professional background of more than five years. Additionally, the verification of human work can be achieved through peer review. Ensuring the validity of human classification, particularly when considering the time investment required for thorough review, necessitates careful attention to selecting suitable individuals. Furthermore, it is advisable to assess or test potential evaluators prior to assigning them classification tasks.

Creating, implementing, and testing static and few-shot text classification

Another classification method was created using the Python programming language and the Pandas library: an “open-source data analysis and manipulation tool built on top of the Python programming language” (Pandas, 2025). The logic checks for the keywords are shown in Figure 2. Additionally, the program ensured that only relevant excerpts of the keywords were identified. For example, for the code Artificial Intelligence and Machine Learning, “AI” would be recognized as a term in the sentence “AI is an important feature.” Conversely, the terms “training” and “gain,” which also contain the keyword “AI,” would not result in a positive result. In summary, this form of evaluation statically checked for existing keywords.

A third classification was performed using a few-shot text classification by GenAI. The term is “particularly pertinent in text classification” and can be defined as “a specialized branch of machine learning, [which] tackles the challenge of constructing accurate models with minimal labelled data” (Aljehani et al., 2025).

However, GenAI must be trained with data before it can be used for classification. Thus, another dataset consisting of 100 sentence examples was created to implement this task, which comprised statements from the study conducted and classified by a human evaluator. Finally, the training and testing data were fed into an instance of ChatGPT5, and the results of the classification were documented and prepared.

Notably, a multi-label rating was possible when a sentence contained multiple technologies simultaneously. For example, the first sentence in Table 2 contains both Artificial Intelligence and Machine Learning, as well as Data Analytics/Science. There was also the option to classify a sentence as “uncategorized” if none of the technologies applied. The results of the classifications appear in Table 3.

A total of 11 of 20 technologies appeared in the test file. GenAI assigned the most codes with 148, followed by the human evaluator (138) and static classification (134).

Table 3: Test dataset classification results

Categories (technologies)	(I) Static classif.	(II) Classif. with GenAI	(III) Human class.
Distr. Ledger Tech. and Blockchain	21	23	25
Data Analytics/Science	15	21	23
AI and ML	21	29	25
Uncategorized	37	17	25

Data/Information Security	12	10	14
Cloud Computing	10	12	10
(Hyper-)Automation	11	13	10
Internet of Things	5	7	5
Digitalization and Digital Transf.	2	11	1
Business Process Mining	0	1	0
Microservices	0	1	0
User Experience	0	3	0
Other categories	0	0	0
Sum of Category Assignments	**134**	**148**	**138**

Evaluating classifications using statistical indicators

Statistical indicators were used to scientifically evaluate the data sets and determine the differences and similarities between the three evaluators. Statistical analysis was performed as part of this research by creating three different evaluation parameters. First was general statistical data, which included easily observable parameters, such as the sum of matching codes, partial matches, and non-matching codes.

Second, the average Jaccard coefficient was calculated by dividing the sum of the individual Jaccard coefficients by the number of sentences. Jaccard index *J* was defined according to Formula 1 below:

$$(1)\ J(A,B) = \frac{|A \cap B|}{|A \cup B|}$$

Here, *J* is calculated by dividing the union of sets A and B by the number of elements that differ between the sets. Consequently, the index values range from 0 to 1. As the index approaches 1, the degree of similarity between the sets increases. Specifically, the value of *J* = 1 indicates perfect congruence, which signifies that the sets are identical.

An illustrative example of the coding process is essential to ensure the index is comprehensible in the context of the article. Table 4 provides this concrete illustration.

Table 4: Example for category assignment

No.	Evaluator 1	Evaluator 2
1	A = {Artificial Intelligence and Machine Learning[1]}	A = {Artificial Intelligence and Machine Learning [1]}, B = {Data Analytics/Science[2]}

In this example, evaluator 1 assigned exactly one technology, while evaluator 2 assigned two. Both agreed on the category Artificial Intelligence and Machine Learning, but Data Analytics/Science was only assigned by evaluator 2.

A single Jaccard index *J* is calculated as follows:

$$(2)\ J(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|\{1\}|}{|\{1,2\}|} = \frac{1}{2} = 0.5$$

Summing up the individual Jaccard indices and dividing them by the number of examples from the test dataset (i.e., 100 in this case) delivers the average Jaccard index.

Third, a more complex metric must be selected that also considers, for example, random agreement between categories. There are several possible agreement statistics, such as Cohen's kappa, Fleiss's kappa, and Krippendorff's alpha.

In this research project, Krippendorff's alpha was chosen as a "standard reliability statistic for content analysis and similar data making efforts" (Krippendorff, 2007). Unlike Cohen's kappa, which is limited to two evaluators, and Fleiss's kappa, which is primarily designed for exclusive categories, Krippendorff's alpha offers the necessary flexibility and robustness to accurately assess agreement under these more complex conditions and substantiate the credibility of qualitative analysis.

Although Krippendorff's alpha is generally unsuitable for multi-label data (Li et al., 2023), the value can be calculated for each category or technology. Generally, the value is calculated by the difference between 1 and the result of the quotient of the observed disagreement D_o and the disagreement expected by chance D_e, as shown in Formula 3:

$$(3)\ \alpha = 1 - \frac{D_O}{D_e}$$

Krippendorff (2011) described D_o as "the observed disagreement among values assigned to units of analysis" and D_e as "the disagreement one would expect when the coding of units is attributable to chance rather than to the properties of these units." In Krippendorff's article referenced in this research, he provided a detailed account of how to compute his index for reliability while elaborating on mathematical methodology and inherent properties. Hence, no further discussion is needed concerning the formula, while reference is made to Krippendorff's original work.

An applicable and straightforward solution for calculating the statistical alpha value was sought within the scope of this research article, so invaluable code libraries were leveraged for this purpose. Krippendorff's alpha was calculated with minimal effort using the Python programming language and a suitable library (Castro, 2017).

The characteristics of each category were translated into a series of binary values to achieve this aim. For example, each sentence to be coded was checked and compared to see whether evaluator 1 and 2 agreed in the category "Data Analytics/Science." This approach allowed the individual categories to be evaluated separately while determining which categories the evaluators agreed on (i.e., a high alpha value) and where there were significant differences of opinion (i.e., a low alpha value).

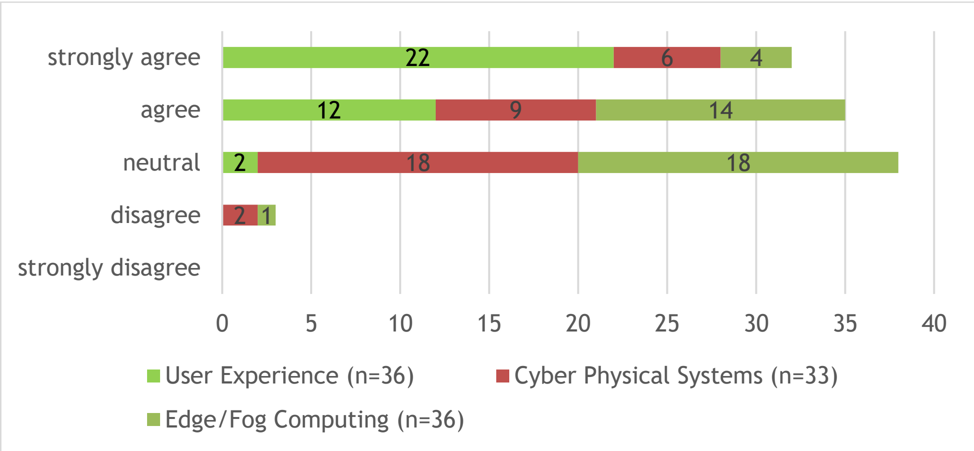
FINDINGS

This section presents the findings from the survey on technological trends in ERP, as well as the evaluation of classifications using statistical indicators. Additionally, a proposal is presented for a new approach to combine GenAI with traditional methods to achieve better and faster results for classifying non-structured text, informed by these findings.

Survey findings concerning technological trends in ERP

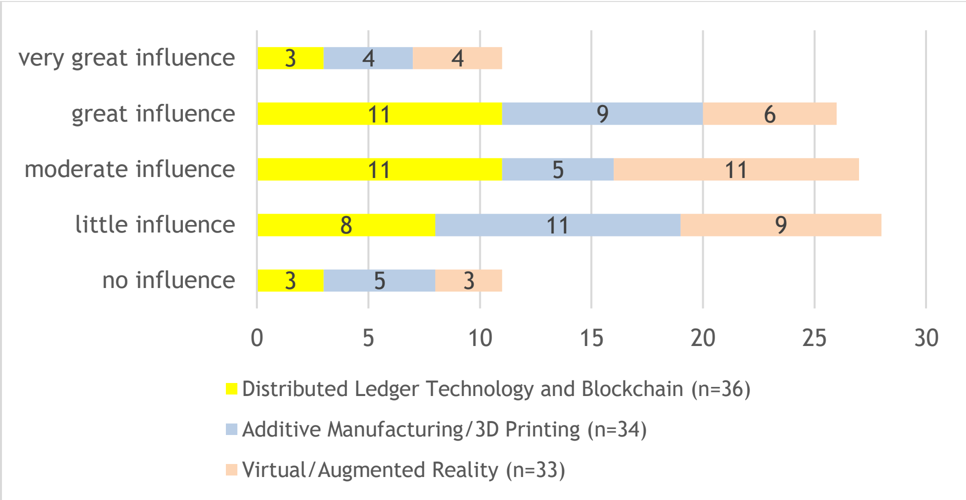
The first technologies the survey participants evaluated were cyber-physical systems and edge/fog computing. User experience, considered in the context of this paper as a technology that utilizes design principles, methods, and tools to design intuitive, efficient, and pleasant interfaces, was also evaluated here. All three technologies were considered to be important for ERP. Their significance is evident in Figure 3, particularly in the strong approval ratings for User Experience, with 22 of the 36 participants (60%) answering this specific question by expressing very strong agreement on its importance. Although approximately one-third of the participants were neutral about the technologies (in the cyber-physical systems category, this figure was as high as 18/33, ~ 55%), only three votes were cast against them.

Figure 3: Importance of user experience, cyber-physical systems, and edge/fog computing (*n* = 105)



Moreover, Figure 4 shows three other technologies: distributed ledger technology and blockchain, additive manufacturing/3D printing, and virtual/augmented reality. Most of the survey participants believed that these three technologies would have a moderate to significant impact on ERP. While approximately 38% (39/103) saw little or even no influence from these technologies, their impact was indeed difficult to deny.

Figure 4: Importance of distributed ledger technology and blockchain, additive manufacturing/3d printing, and virtual/augmented reality (n = 103)

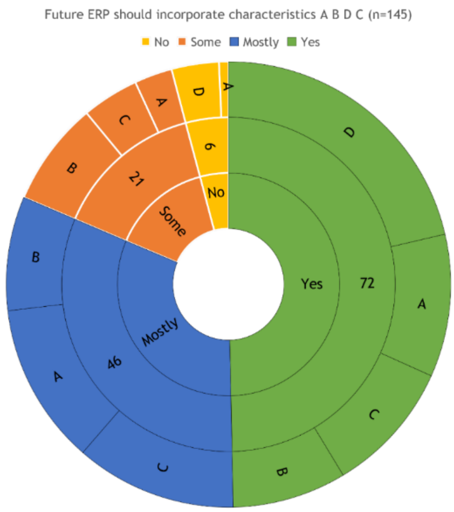


Finally, Figure 5 shows four desirable characteristics of ERP systems from the participants' perspectives (i.e., A to D), which reflect characteristics from four technological areas: (A) Microservices, (B) In-Memory Computing, (C) Business Process Mining, and (D) Low/No-Code. Approximately 81% (118/145) of the respondents would consider an ERP system with these four characteristics, depicted in Table 5, to be useful. Conversely, only 21 respondents considered these attributes important to some degree, while only six deemed them entirely unimportant.

Table 5: Requirements for ERP systems

Req.	Description
A	The future architecture of ERP software should be based on innovative cloud architecture with microservices and open APIs.
B	ERP systems should use the latest technology (e.g., fast in-memory computing) with the ability to process large volumes of data instantly.
C	ERP should offer more extensive analysis and monitoring functions to exploit potential for process improvements, which can be implemented directly in the ERP system or via preconfigured connectors for process mining tools.
D	ERP should provide stable interfaces to external NC/LC platforms so that the ERP system remains the central instance of data sovereignty.

Figure 5: Desirable requirements for ERP systems according to the survey participants



The survey included an optional free-text question that asked the respondents about their opinions on technologies affecting ERP that were not previously mentioned in this article. Many answers included AI. Moreover, 5G and quantum computing were addressed three times each, which suggested that these technologies were highly regarded by the experts and relevant to the classification scheme.

Findings from evaluating classifications using statistical indicators

Table 6 presents various statistical indicators, which are explained below. They were useful for evaluating the implemented classification options. The key figures refer to the data set of the human evaluator, which serves as a benchmark.

Table 6: Statistical indicators

Statistical Indicators	(I) Static Classification	(II) Classif. with GenAI
General Data		
Matching codes	77	61
Partial matches	9	25
Non-matching codes	14	16
Average Jaccard Index	0.81	0.73
Krippendorff's Alpha		
Cloud Computing	1.0	0.90
Distr. Ledger Tech. and Blockchain	0.89	0.95
AI and ML	0.89	0.85
Digitalization and Digital Transf.	1.0	0.12

Internet of Things	1.0	0.82
Data/Information Security	0.91	0.72
Data Analytics/Science	0.75	0.69
(Hyper-)Automation	0.74	0.66
Uncategorized	0.67	0.64
Business Process Mining	-	0.0
Microservices	-	0.0
User Experience	-	0.0

First, general data is displayed, which shows the sum of matching codes, partial matches, and non-matching codes. Notably, the static classification corresponds to 77% (77/100) of the authors' classification, a very good result, whereas GenAI achieved only 61% (61/100). In contrast, there were nine sentences in which the codes of the static classifier partially matched, compared to 25 for GenAI. There was complete disagreement in 16 cases with AI, and 14 cases with static classification. Therefore, the general statistical indicators showed that the static classification provided more absolute matching codes (77 vs. 61) and fewer absolute non-matching codes (14 vs. 16) than GenAI.

Furthermore, a comparison of the average Jaccard index of the two data sets also yielded a better value of 0.81 for static classification than 0.73 for classification with ChatGPT. The closer the value was to 1, the higher the agreement between the data sets. When considering all text passages, the codes assigned by the Python script overlapped 81% on average with those of the human evaluator, based on the total number of unique labels.

Nonetheless, Krippendorff's alpha values in the individual categories provided a differentiated picture. Krippendorff recommended "rely[ing] only on variables with reliabilities above *a* = .800" while "consider[ing] variables with reliabilities between *a* = .667 and *a* = .800 only for drawing tentative conclusions" (Krippendorff, 2004). Using these thresholds as an approximate guide, some notable differences between the two evaluators (i.e., static classification and classification with GenAI) were seen.

The results of the static classification indicated that the "uncategorized" category was only slightly above the threshold of 0.667 due to the insufficient performance of the Python script. In Table 7, Example 2 illustrates that static classification only recognized the string "pattern recognition," which would have resulted in classification in the category "Data Analytics/Science" but not the term "patterns." Example 9 showed a similar result, where "analysis of these vast datasets" was not recognized as a string. However, a human evaluator familiar with the technologies should be able to classify one easily.

Table 7: Wrong example category – Uncategorized

Nr.	Sentence	(I) Static	(III) Human
2	"These sophisticated methods can analyze complex datasets and uncover *patterns* that traditional	Uncategorized	Data Analytics/ Science

	statistical techniques might miss, resulting in more reliable forecasts" (Mhaskey 2025: 6).		
9	"They deliver extensive storage options and robust computing resources, allowing for the efficient processing and *analysis of these vast datasets*—capabilities often unattainable for many small and mid-sized businesses" (Mhaskey 2025: 6).	Uncategorized	Data Analytics/ Science

The results of classification with GenAI were particularly interesting, as some categories stood out clearly. One anomaly was the very poor score of 0.12 for Digitalization and Digital Transformation. According to the number of classifications in Table 3, GenAI assigned this category 11 times. Compared to (I) and (III), this deviation was significant. Moreover, examining sentence 20 in Table 8, GenAI might interpret the string "transformative effect" as referring to the category Digitalization and Digital Transformation. In addition, example 48 suggests that GenAI assumes institutional change also requires digital components.

Table 8: Wrong example category – Digitalization and Digital Transformation

Nr.	Sentence	(II) GenAI	(III) Human
20	"Exploring ERP analytics has illuminated its *transformative effect* on organizational performance, facilitating enhanced decision-making, operational efficiency, and strategic financial management" (Mhaskey 2025: 7).	Data Analytics/Science, Digitalization and Digital Transformation	Data Analytics /Science
48	"Current predictions also suggest that the future path of ERPs is a movement towards adaptation to revolutionize service co-production or value co-creation, necessitating organizations and institutional change" (Kaulwar 2025: 33).	Digitalization and Digital Transformation	Uncategorized

Observing the alpha values for the categories Business Process Mining, Microservices, and User Experience revealed that AI consistently failed to classify them correctly, which resulted in alpha values of 0.0. Although these were only assigned a total of five times (see Table 3), two examples are provided in Table 9. Sentence 74 revolves around "organizing workflows efficiently," which GenAI may associate with business processes. The second example, sentence 75, describes how a "high-quality degree of service" satisfies the customer. In this case, one could certainly speak of a good "user experience" if the description refers to the interaction with the ERP system. However, since this idea could only be assumed, the category shows a difference.

Table 9: Wrong example categories – Business Process Mining and User Experience

Nr.	Sentence	(I) Static C	(III) Human

74	“*Organizing workflows efficiently* can result in rapid or almost immediate real-time decisions and therefore quick responses” (Kaulwar 2025: 34).	(Hyper-) Automation, Business Process Mining	Uncategorized
75	“This operational speed can be transposed to customers and internal clients who could benefit from an immediate and high-quality degree of service” (Kaulwar 2025: 34).	User Experience	Uncategorized

Two categories, (Hyper-)Automation (0.66) and Uncategorized (0.64), also fell below the threshold when evaluated using Krippendorff’s alpha, so they had to be adjusted. However, the analysis was similar to the previous ones, so further discussion is unwarranted here.

Finally, another striking pattern was noteworthy. In some sentences, it was debatable whether the AI correctly recognized the context of the sentences. Two of these examples are shown in Table 10. Example 26 refers to the fact that the implementation of “these solutions” (i.e., automation and AI) calls for certain requirements. In Example 57, “these innovations” refer to AI and Big Data Analytics, mentioned in the previous sentence. These examples illustrate the challenges that non-human evaluators face in this context.

Table 10: Wrong example – Missing context

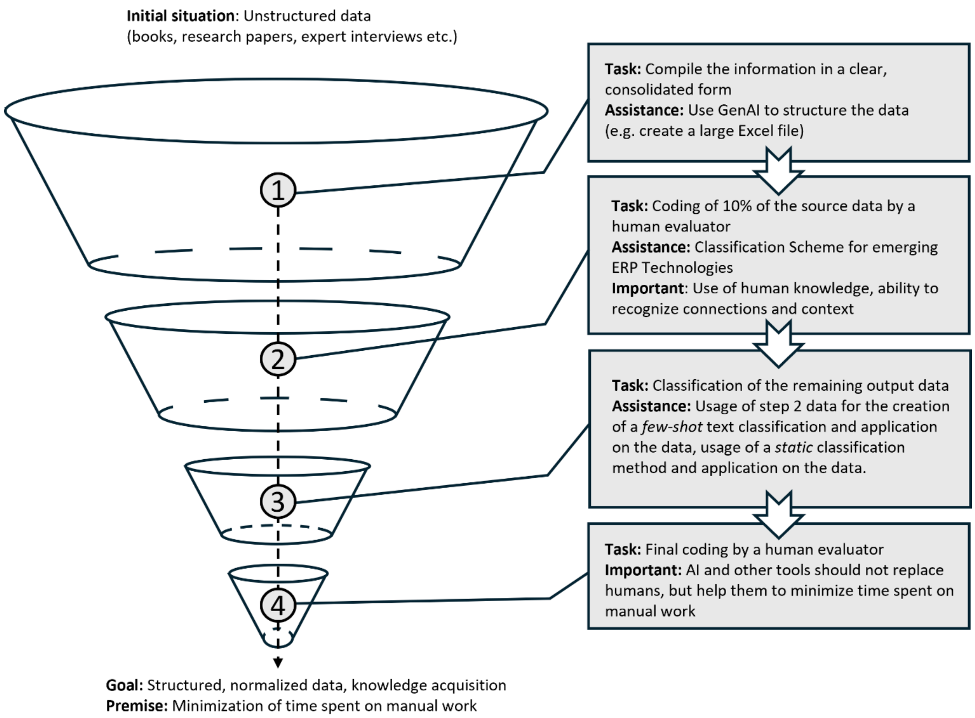
Nr.	Sentence	(II) GenAI	(III) Human
6	“The successful implementation *of these solutions* requires careful consideration of both technical integration requirements and organizational change management strategies” (Onteddu 2025: 596).	Digitalization and Digital Transformation	Artificial Intelligence, (Hyper-)Automation
57	“*These innovations* will change the perspective of ERP systems considerably” (Kaulwar 2025: 33).	Digitalization and Digital Transformation	Artificial Intelligence and Machine Learning, Data Analytics/Science

Proposed framework for efficiently classifying unstructured text

This scientific work proposes a novel approach that combines GenAI with traditional methods for achieving better and faster results in classifying non-structured texts. Figure 6 presents a four-step funnel-based framework for efficiently classifying unstructured text. It presents recommendations derived from the findings of this research and outlines how to obtain structured, normalized data and acquire knowledge in four steps from a large amount of unstructured data (e.g., books, research papers, and expert interviews).

This approach can significantly reduce the time spent on manual work while maintaining the best accuracy achieved through classification by a domain expert. The steps are presented in the form of an analogy to a funnel. Just as a funnel diverts a liquid into a controlled stream by means of ever-increasing narrowing, the proposed process is intended to convert a large amount of unstructured data into usable information and deliver structured, normalized data.

Figure 6: Four-step funnel-based framework for efficient classification of unstructured text



The first recommended task is to organize the material into a structured, consolidated form. GenAI is used to structure the source material. In this work, whole sentences were chosen as the smallest classifiable units. As a result, the unstructured source material is presented in a table in which the quotations are listed sequentially.

Secondly, at least 10% of the source material should be classified by a human evaluator. The classification scheme for emerging ERP technologies, illustrated in Figure 2, should be used here. This classification guidance can be expanded and adapted to new keywords during a research project.

The third step can involve either the static classification method or a few-shot text classification approach. In the case of the static classification method, the logic must then be expanded to include any new keywords that may have been created in Step 2. In the case of few-shot text classification, the data used in Step 2 can be utilized as training data for GenAI. The result is a table with classified units, which may resemble the excerpts in Tables 7–10. Notably, these results only correspond to those of a human evaluator to a limited extent.

As a final step, this table must be reviewed by a competent researcher. Although this work presents arguments for a correlation between the tools and a “real” human evaluator that can at least be substantiated by statistical indicators, these tools cannot replace or even substitute for human researchers. However, using this approach is based on a solid framework and can save considerable additional effort.

CONCLUSIONS AND FURTHER RESEARCH

The classification task for determining the categories of new technologies that will affect the functionality and future development of ERP systems is of the unstructured text processing type. The best results can be achieved with manual classification by a specialist in the field, but the task is time-consuming. Increasing efficiency can be achieved through various approaches for natural language automation using AI techniques, which involves building a knowledge base with code words in a specific domain for training a classification model and then continuously improving it for multiple uses within the same discipline. This approach is effective when processing text data in a specific field. The bases are used repeatedly and enriched with each new set of data.

The task of the present study was classification in a new domain for which there was no established baseline. Hence, a new approach is proposed to achieve a qualitative and rapid assessment of technologies that impact the development of ERP systems. By training GenAI in the new domain and optimizing the classification process, this innovative approach follows a few consecutive steps. The classification framework of the approach describes in detail the stages of processing unstructured text. GenAI is applied to approximately define the main categories and significantly reduce the classification time. The data is verified by static logic and refined by a domain expert. The proposed approach preserves the quality of the classification result while increasing the operational speed. The classification framework is suitable for processing unstructured data in a new domain for which there is no existing code base, yet fast and high-quality analysis is required.

The method developed in the study can be expanded in future work to larger and more diverse source materials (e.g., detailed expert interviews) to further demonstrate its practicality. Additional surveys and literature analyses in the field of ERP may reveal recently unknown ERP trends and developments that must be integrated into the process.

ACKNOWLEDGMENTS

The paper is realized within the project number NIP-2025-11, “Innovation through data-driven IT applications and advanced analytics,” funded by the Ministry of Education and Science, Bulgaria.

REFERENCES

Aljehani, A., Hamid Hasan, S., & Ali Khan, U. (2025). Advancing text classification: A systematic review of few-shot learning approaches. *International Journal of Computing and Digital Systems, 17*(1), 1–21. https://doi.org/10.12785/ijcds/1571027526

Castro, S. (2017). Fast Krippendorff: Fast computation of Krippendorff's alpha agreement measure. GitHub repository. https://github.com/pln-fing-udelar/fast-krippendorff

Kaulwar, P. (2025). Enhancing ERP systems with big data analytics and AI-driven cybersecurity mechanisms. *Journal of Artificial Intelligence and Big Data Disciplines, 2*(1), 27–35. https://doi.org/10.70179/32w32914

Krippendorff, K. (2011). Computing Krippendorff's Alpha-Reliability. https://repository.upenn.edu/asc_papers/43

Li, S., Yan, R., Wang, Q., Zeng, J., Zhu, X., Liu, Y., & Li, H. (2023). Annotation quality measurement in multi-label annotations. In F. Liu, N. Duan, Q. Xu, & Y. Hong (Eds.), *Natural Language Processing and Chinese Computing: 12th National CCF Conference, NLPCC 2023, Foshan, China, October 12–15, 2023, Proceedings, Part II, Lecture Notes in Computer Science* (pp. 30–42). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-44696-2_3

Mhaskey, S. (2025). Unlocking business potential: The transformative impact of ERP analytics. *Journal of Advances in Developmental Research, 16*(1), 1–9. https://doi.org/10.5281/zenodo.15155165

Onteddu, K. (2025). AI-powered invoice automation in ERP systems: Revolutionizing accounts payable. Journal of Computer Science and Technology Studies, 7(5), 588–597. https://doi.org/10.32996/jcsts.2025.7.5.64

Pandas. (2025). *Pandas documentation*. https://pandas.pydata.org/

Sunmola, F., & Lawrence, G. (2024). Key success factors for integration of blockchain and ERP systems: A systematic literature review. *Procedia Computer Science, 232*, 775–782. https://doi.org/10.1016/j.procs.2024.01.077