International Journal of Social, Political and Financial Researches, 2025, 5(2), 302-313
DOI: 10.70101/ussmad.1758962

Uluslararası Sosyal Siyasal ve Mali Araştırmalar Dergisi

International Journal of Social, Political and Financial Researches

https://dergipark.org.tr/tr/pub/ussmad

Araştırma Makalesi/ Research Article

Using Generative Artificial Intelligence as a Decision Support Tool in Purchasing Processes: Comparison of ChatGPT, CoPilot, and Gemini Tools

Üretken Yapay Zekânın Satın Alma Süreçlerinde Karar Destek Aracı Olarak Kullanılması: ChatGPT, CoPilot ve Gemini Araçlarının Karşılaştırılması

Hakan Aşan[a]
[a]Asst. Prof. Dr., Dokuz Eylul University, hakan.asan@deu.edu.tr, ORCID: 0000-0001-9550-3345

ARTICLE INFO

Article Received: 06.08.2025

Article Accepted: 06.10.2025

Keywords: Generative Artificial Intelligence, Purchasing Processes, Decision Support System, Enterprise Resource Planning (ERP)

JEL Codes: C45, M15, O33

ABSTRACT

Accurate prioritization of purchase requests in enterprises is critical for ensuring business continuity and effective resource management. Throughout the day, requests generated by different departments are usually ranked subjectively by the purchasing unit, which may cause some urgent requests to be deprioritized. Managing the process under human control leads to time loss and inaccurate prioritization. This study integrated three generative artificial intelligence tools—ChatGPT-4.5, Microsoft CoPilot, and Google Gemini—into a manufacturing company's ERP system via an API. A total of 100 purchase requests were classified first into three categories ("Urgent," "Normal," and "Not Urgent") and then into two categories ("Urgent" and "Normal"). The results produced by the AI models were compared with the classifications made by the purchasing staff and evaluated using accuracy, Cohen's Kappa, precision, recall, and F1-score metrics. In addition, the correct response performance of generative artificial intelligence tools was analyzed using the Pearson Chi-square test; the results revealed a significant interdependence among the tools, with Copilot and Gemini showing an exceptionally high consistency across both triple and binary classifications. The findings revealed that all three models performed well in the binary classification, with CoPilot achieving higher accuracy than the others. The study demonstrates that generative AI tools can be practical decision-support systems in purchasing processes, offering significant advantages in preliminary classification, efficiency, and time savings.

MAKALE BİLGİSİ

Makale Gönderim Tarihi: 06.08.2025

Makale Kabul Tarihi: 06.10.2025

Anahtar Kelimeler: Üretken Yapay Zeka, Satın Alma Süreci, Karar Destek Sistemi, Kurumsal Kaynak Planlama (KKP)

JEL Kodları: C45, M15, O33

ÖZ

İşletmelerde satın alma taleplerinin doğru önceliklendirilmesi, iş sürekliliği ve kaynak yönetimi açısından kritik öneme sahiptir. Gün içinde farklı departmanlarca oluşturulan talepler genellikle satın alma birimi tarafından sübjektif olarak sıralanmakta, bu da bazı acil taleplerin geri planda kalmasına yol açabilmektedir. Sürecin insan kontrolünde yürütülmesi hem zaman kaybına hem de hatalı önceliklendirmelere neden olmaktadır. Bu çalışmada, bir üretim işletmesinin ERP sistemine API aracılığıyla entegre edilen üç üretken yapay zekâ aracı (ChatGPT-4.5, Microsoft CoPilot ve Google Gemini) kullanılarak 100 satın alma talebi önce "Acil", "Normal" ve "Acil Değil", ardından "Acil" ve "Normal" biçiminde sınıflandırılmıştır. Yapay zekâ modellerinin sonuçları, satın alma personelinin sınıflandırmalarıyla karşılaştırılarak doğruluk, Cohen's Kappa, precision, recall ve F1-score metrikleri üzerinden değerlendirilmiştir. Ayrıca üretken yapay zeka araçlarının doğru yanıt verme performansları Pearson Ki-kare testiyle incelenmiş; sonuçlar, araçlar arasında anlamlı bir karşılıklı bağımlılık olduğunu ve özellikle Copilot ile Gemini'nin hem üçlü hem de ikili sınıflandırmalarda yüksek düzeyde uyum sergilediğini göstermiştir. Bulgular, her üç modelin ikili sınıflandırmada başarılı performans gösterdiğini; özellikle CoPilot'un diğer modellere göre daha yüksek doğruluk sağladığını ortaya koymuştur. Çalışma, üretken yapay zekâ araçlarının satın alma süreçlerinde ön sınıflandırma, hız ve iş gücü tasarrufu açısından etkili bir karar destek aracı olarak kullanılabileceğini göstermektedir.

Introduction

The role of purchasing departments in businesses is not limited to price and delivery negotiations with suppliers. It also involves making purchasing decisions, balancing warehouse and product supply, and, most importantly, determining the purchasing order. Sequencing purchasing requests is crucial, especially for businesses operating 24/7 for production. Production interruptions, especially when orders for products requiring long procurement times are not placed at the right time, can make enterprises incapable of managing their business processes. The process begins with a purchase request and is driven entirely by subjective decisions made by experienced purchasing personnel. However, the multifaceted nature of the purchasing process affects the quality of the decisions made and can cause problems for the business. In this sense, using artificial intelligence as a decision support system at specific points will enable businesses to manage their purchasing processes more effectively and efficiently. Generative AI tools, which are becoming increasingly common today, can support companies in decision-making without requiring technical knowledge. Unlike previous digital and industrial revolutions, generative AI tools are poised to perform all tasks rather than just a few (Makridakis, 2017). Huang et al. (2022) say that generative AI can improve the efficiency of the knowledge and creativity workforce by at least 10%, making them more efficient and skilled. The market share of generative artificial intelligence, estimated at $13.71 billion in 2023, is expected to exceed $100 billion in 2032 (Precedence Research, 2025). This study investigates whether generative AI tools can be practical support tools for businesses' purchasing processes. Specifically, it aims to evaluate the individual and the interplay of generative AI tools in purchasing, a crucial step for manufacturing businesses. To this end, the study categorized purchase requests entered via API connections to generative AI tools within a manufacturing company's enterprise resource planning (ERP) software into three categories: "Urgent," "Normal," and "Not Urgent." These requests were further categorized into "Urgent" and "Normal". Expert personnel also classified and recorded the first 100 requests entered into the ERP system. The study aimed to evaluate the performance of artificial intelligence tools in the request prioritization process and to test their usability as a decision support mechanism. Evaluations were conducted using accuracy, Cohen's Kappa, precision, recall, and F1-score metrics. In this respect, the study provides a theoretical contribution to integrating artificial intelligence into supply chain processes and a guiding framework for practical applications.

In the first phase of the study, generative AI tools were introduced. The second section included studies on this topic and presented their evaluations. The third section outlined how and where the data were obtained, and the fourth section presented the findings. The conclusion section provided a comparative analysis of the findings, limitations, and recommendations for future research.

1. Generative Artificial Intelligence Applications

The concept of artificial intelligence was first expressed by John McCarthy in 1955 (McCarthy et al., 1955). Since its formulation, McCarthy had developed the concept, and several scientists laid the first foundations of artificial intelligence at the Dartmouth Conference held in 1956 (Moor, 2006, p. 87). Artificial intelligence, defined, is the imitation of human intelligence by machines. In other words, it is a collection of technologies capable of performing human-like activities such as thinking and learning. One of the key capabilities of artificial intelligence is processing and understanding natural language. Owing to this ability, the current state of large language model algorithms is remarkable. The development of these technologies has given rise to a new concept. This new concept, generative artificial intelligence, can generate various data types, such as text, images, audio, and synthetic data (Lawton, 2023). The original version of the concept, called "Generative Pretrained Transformer" (in English), is Generative Pretrained Transformer (GPT) (Zhu et al., 2023). Generative AI tools have quickly attracted significant interest from end users. ChatGPT (Chat Generative Pretrained Transformer), launched by OpenAI in November 2022, reached 1 million users less than five days after its launch. By July 2023, this number had surpassed 100 million, and monthly visits reached 1 billion (Duarte, 2025; Thormnundsson, 2023). One of the most important features of generative AI is its ability to produce various types of content. Previously used tools successfully processed and classified text, images, and sounds. However, generative AI tools also hold the potential to create new things (Wiles, 2023). This potential allows for quickly completing tasks that might otherwise take a long time, freeing up time for other tasks (Deloitte AI Institute, 2023, p.11). Generative AI tools are frequently used in fields such as law, sports, tourism, the environment, biology, mathematics, and health, as well as for topics such as article writing, news review, and website updating (Göktaş, 2023, p. 894; Mich & Garigliano, 2023, p. 4).

The most popular generative AI tool is ChatGPT (Chat Generative Pretrained Transformer), released by OpenAI in November 2022. ChatGPT quickly gained widespread acceptance. AI software like ChatGPT, introduced as the latest machine learning technology with the GPT-3.5 language model, currently demonstrates its ability to pass the Turing Test (Pavlik, 2023). OpenAI developed the GPT-1, GPT-2, GPT-3, and GPT-4 language models (Koubaa, 2023). Other popular examples of generative AI include text generators like Google Bard and Bing AI,

visual generators like DALL-E and Midjourney, music generators like Ampere and MuseNet, and code generators like CodeStarter, Codex, and GitHub CoPilot (Lawton, 2023). Due to OpenAI's success and end-user impact, Google released its generative AI tool, Gemini, in March 2023. Gemini is designed to operate in a multimodal capacity by facilitating processing various data types, such as text, images, audio, and video. Gemini's core architecture incorporates elements from Google's previous models, including the Language Model for Conversational Applications (LaMDA) and the Pathways Language Model (PaLM) (Rane et al., 2024; Rossettini et al., 2024; Gupta et al., 2024). Microsoft, one of the industry's leading companies, developed another generative AI tool. This tool, called CoPilot, is part of next-generation AI solutions that increase human productivity in various areas, especially programming and office productivity (Zhang et al., 2023; Wong et al., 2023; Chatterjee et al., 2024).

Despite the positive features of generative AI tools, some ethical concerns should also be considered. Serious concerns exist, particularly in vital fields such as the healthcare sector. When addressing complex medical questions, concerns arise about the accuracy and reliability of the responses provided by systems or individuals (Sallam, 2023; Dave et al., 2023). Due to the lack of a physical examination and complex assessments, it is not yet considered sufficient for evaluating real-life cases (Dabbas et al., 2024). Therefore, it is considered more appropriate to use it for educational purposes or as a complementary tool in such cases (Li & Guenier, 2024). Furthermore, because generative AI is a data-driven system, it carries some concerns due to embedded biases and potential misuse (Dis et al., 2023; Ventura & Filho, 2023). However, this concern is expected to decrease with increasing data accuracy (Al-Worafi et al., 2023). Another concern is its generative capacity. Generative AI tools can generate inaccurate data when the data is insufficient or based on incorrect references. For example, when considering an academic text, there is no definitive information about the accuracy of the sources. Such studies raise concerns (Liu et al., 2023; Osama & Afridi, 2023).

2. Literature

Numerous studies on the use of artificial intelligence tools exist in the literature. These studies are designed to measure generative artificial intelligence's effectiveness and compare its tools.

A growing body of research has examined the effectiveness of generative AI applications in education, training, and exam evaluations. Jalil et al. (2023) asked ChatGPT questions from a book on an undergraduate software testing course and determined its ability to answer the questions. They determined that ChatGPT answered 77.5% of the questions, with 55.6% correct answers. Küçüker (2023) tested ChatGPT's knowledge adequacy in basic accounting topics and demonstrated the model's potential contributions, particularly in financial accounting, cost-management accounting, and auditing. Boduroğlu et al. (2023) compared ChatGPT's ability to classify the difficulty levels of multiple-choice test items with expert opinions and obtained high correlation values. Yalçın-Çelik and Çoban (2023) analyzed the accuracy of responses to university-level chemistry questions based on Bloom's cognitive domain taxonomy. Kutlucan and Seferoğlu (2024) evaluated the impact of ChatGPT, a generative AI tool, on learning and teaching processes. The study categorized 150 articles according to KEFE and PEST analyses. While the study indicated that generative AI tools could provide equal opportunities in education, the tools also raised plagiarism and ethical concerns. Harada et al. (2025) measured the impact of data volume on accuracy in automatic book classification by training a GPT-3.5 model with data from the national library in Japan. Çelik et al. (2025) evaluated the success of generative AI in multiple-choice questions based on language differences. The differences between questions presented in English and Turkish were demonstrated. Emekli et al. (2025) shared their experiences with scenario creation and multiple-choice question generation using the ChatGPT-4.0 generative AI tool in medical psychiatry education, emphasizing the potential of this process to support students' clinical reasoning skills. A significant level of consistency was achieved when scoring criteria were defined. Another area of research is focused on integrating generative AI tools into accounting and finance processes. Ahmadi (2023) evaluated the potential role of OpenAI solutions against fraudulent activities in the financial sector. Smales (2023) evaluated the classification of central bank statements as _hawk_ or _dove_ through ChatGPT. Murindanyi et al. (2023) used ChatGPT and Explainable AI methods to predict customer interaction in the banking sector. Accuracy rates of up to 99% were achieved in the study. Studies in the education sector show that generative artificial intelligence tools provide high accuracy. The healthcare sector is another area where generative AI technologies are rapidly being adopted and used. Das et al. (2023) asked ChatGPT questions requiring high-level thinking and interpretation related to microbiology. The study found that the questions were answered correctly, with an accuracy rate of 80%. (Dökme Yağar, 2023) evaluated the potential of generative AI in healthcare using ChatGPT as an example, considering ethical concerns and legal requirements. Liu et al. (2024) demonstrated the high accuracy and recommendation capability of GPT-4.0 in the automatic construction of breast ultrasound reports. Guo & Wan (2024) determined that multimodal LLMs can produce more accurate results than traditional methods in classifying patient medical images. Öztürk and Ergin (2024) used generative AI to evaluate ChatGPT's ability to answer questions about

lower urinary tract symptoms in women. Shaheen et al. (2025) used the generative AI tool ChatGPT to classify bleeding after eye surgery. High accuracy rates were achieved in the study. Brigo et al. (2025) evaluated the performance of ChatGPT-4.0 in diagnosing patients with epileptic seizures and classifying diseases. Comparing the results with those of generative AI and the field experts' responses, they noted that the model achieved high sensitivity but lacked specificity. In the healthcare sector, generative artificial intelligence is used in studies with various characteristics and achieves high accuracy rates.

Other studies in various fields evaluate the use and effectiveness of generative AI tools. Sudirjo et al. (2023) analyzed the effectiveness of ChatGPT in analyzing customer feedback. Dalgıç (2023) analyzed whether generative AI could manage and resolve restaurant customer complaints. Banimelhem and Amayreh (2023) analyzed the DAIR.AI dataset using ChatGPT for sentiment classification. The study reported a 58% accuracy rate. Çeber (2024) evaluated the use of ChatGPT and Midjourney tools in advertising agencies. Erdem (2024b) analyzed the websites of logistics companies operating in Turkey using generative AI based on 11 criteria. Küçük and Can (2024) discussed the potential of generative AI tools like ChatGPT for automated processing of legal documents, highlighting these systems' fast, accurate, and standardized document processing capabilities. Uyar (2024) evaluated the GPT-3.5 model's ability to detect logical fallacies in his study. The study revealed that the method could distinguish fallacy types with high accuracy. Darlan et al. (2025) studied generative artificial intelligence in agriculture. AI-supported image recognition systems were used to classify the growth stages of strawberry fruit.

Another method regarding generative AI tools is to compare them. Studies similar to this one evaluate the effectiveness of the tools and their comparison with each other. Butean et al. (2025) compared different ChatGPT models for evaluating cases on a specific topic in dentistry. The study claims the GPT-4 Pro model will offer significant clinical contributions. Erdem (2024a) examined the usability of three different AI tools (ChatGPT, Bing, and YouChat) in the social sciences. The study evaluated ChatGPT and Bing as competent in all categories, while YouChat was found to have more limited competence. Rane et al. (2024) compared ChatGPT and Gemini generative AI models. The evaluation was based on data sensitivity, ease of integration, cost, and intended use. Şensoy and Çıtırık (2025) answered 32 questions about glaucoma using ChatGPT-3.5, Gemini, and CoPilot. ChatGPT-3.5 answered 68.8% of the questions, Gemini answered 43.8%, and CoPilot answered 56.3%. ChatGPT-3.5, Gemini, and CoPilot answered the Turkish questions correctly at 34.4%, 36.5%, and 34.4%, respectively. ChatGPT-3.5 answered English questions significantly more successfully than it did Turkish questions. No difference was found between Gemini and CoPilot AI chatbots in correctly answering English and Turkish questions. ChatGPT 3.5 answered 45% (n=18) of the questions, ChatGPT 4.0 answered 52.5% (n=21) of the questions, Gemini answered 87.5% (n=35) of the questions, and CoPilot answered 60% (n=24) of the questions correctly. Yılmaz and Çil (2024) investigated the answers to 80 questions about erectile dysfunction. The study evaluated Google AI and ChatGPT responses. Both tools provided accurate and satisfactory answers to questions about erectile dysfunction, but performance fluctuated between questions. Ayhan and Kılıç (2024) conducted a detailed analysis of the responses provided by ChatGPT and Gemini to eight questions on the history of hadith. ChatGPT offered a more systematic and in-depth approach to the topic, while Gemini provided a more superficial and general perspective. Okur and Ekşi (2024) examined the reliability and understandability of the responses of ChatGPT and Gemini, an artificial intelligence model developed by Google AI, also based in the US, to questions posed to theology students regarding property within the context of Islamic Property Law. In the study, both models were asked questions at easy, medium, and challenging levels, and their ability to present and analyze information on general legal concepts, fundamental principles, and conceptual analyses was evaluated. Gelmiş et al. (2025) compared the performance of ChatGPT-4 and Google Gemini in patient education regarding penile prosthesis use. The study collected 50 questions from the "People also ask" section of Google search results. Two experienced urologists independently evaluated the responses using the Global Quality Score (GQS). ChatGPT-4 outperformed Google Gemini by providing both faster and more accurate responses.

3. Data

The data for this study were obtained from purchase requisitions entered through the ERP system of a three-shift manufacturing organization. Requisitions were recorded instantaneously using generative AI APIs. The purchasing personnel and three generative AI tools (ChatGPT-4.5, Gemini, and CoPilot) classified the collected data as “urgent,” “normal,” and “not urgent.” The classifications of the AI tools were compared with those of the procurement staff.

Individuals or organizations must provide accurate and meaningful data for a specific job to fully exploit generative AI's potential (Brynjolfsson et al., 2023, pp. 7-8). Therefore, the same custom prompt was used for all AI tools in the study. The prompt was prepared by considering the following criteria:

- Is the requested material production-related?

- Are there time constraints?
- What is the type of material (e.g., machine, bearing, raw material, office product)?
- Is it related to a breakdown or repair?
- Was the request made by management?
- Is the request a periodic purchase (e.g., contracts, consultancies)?

4. Findings

The 100 purchase requests in the study were subjected to a triple classification as “urgent”, ‘normal’ and “not urgent” by three artificial intelligence models. The purchasing personnel also classified the exact requests through ERP. In addition, binary classification into "urgent" and "normal" was also performed with the same data.

Table 1: Confusion Matrix for the Three Models

Model	Actual \ Estimate	Urgent	Normal	Not Urgent
	Urgent	14	13	6
ChatGPT	Normal	8	32	8
	Not Urgent	0	12	7
	Urgent	17	15	1
CoPilot	Normal	3	41	4
	Not Urgent	1	15	3
	Urgent	14	19	0
Gemini	Normal	3	35	10
	Not Urgent	0	7	12

According to Table 1, the success of the models in recognizing the "normal" class is remarkable. Of the 45 "normal" requests identified by the procurement staff, ChatGPT correctly classified 32 (71%), CoPilot 41 (91%), and Gemini 35 (77%). In terms of overall accuracy, CoPilot stands out as the most successful model; however, it needs improvement in the "not urgent" class. Gemini is the most successful model in this class. ChatGPT, conversely, exhibits significant confusion, especially with the "not urgent" class, and has difficulty distinguishing between classes in general.

Based on the confusion matrix data, the metrics of the triple classification are presented in Table 2.

Table 2: Three-Class Classification Metrics

	Accuracy	Precision	Recall	F1-Score	Cohen's Kappa
ChatGPT	0.53	0.76	0.45	0.57	0.23
CoPilot	0.61	0.57	0.75	0.65	0.32
Gemini	0.61	0.42	0.39	0.39	0.36

The results reported in Table 2 indicate that CoPilot has the highest accuracy of 61% and exhibits balanced classification performance. ChatGPT can be considered a cautious model choice that avoids misclassifications. Gemini is the least preferred model due to its low Precision and Recall values. Due to the low metrics obtained in the triple classification, the analysis was repeated using the "normal" and "urgent" binary classification currently adopted by the enterprise. The same 100 demands were reclassified with three models. To eliminate the effect of the previous classification, the memory of the models was reset, and the prompt was reorganized. The results obtained are presented in Table 3.

Table 3: Binary Classification Metrics

	Accuracy	Precision	Recall	F1-Score	Cohen's Kappa
ChatGPT	0.81	0.81	0.71	0.74	0.48
CoPilot	0.83	0.82	0.75	0.78	0.56
Gemini	0.81	0.81	0.71	0.74	0.48

The results reported in Table 3 indicate that the CoPilot model has the highest performance in all metrics. It is the most reliable and consistent model with 83% accuracy and a Cohen's Kappa value of 0.56. The high Recall and F1-score values indicate that the model accurately captures many positive examples. ChatGPT and Gemini produced similar results with 81% accuracy and a Cohen's Kappa value of 0.48. A Cohen's Kappa value between 0.41 and 0.60 is considered a moderate level of agreement (Crewson, 2005). Based on the binary classification metrics, it can be seen that all three models classify successfully.

The correct and incorrect answers generated by AI tools were tested for independence using the Pearson Chi-square test at both three-fold and two-fold classification levels. The tests were conducted using the IBM SPSS 25 software package.

When evaluating the three-fold classification results, a significant correlation was found between the correct and incorrect answers provided by ChatGPT and CoPilot for 100 questions ($\chi^2 = 9.927$, $p = 0.002$). This indicates that the responses from the two AI systems are not independent and display similar patterns of correct or incorrect answers. In other words, when one system provides a correct answer, the probability that the other system will also provide a correct answer increases significantly. However, no statistically significant difference was observed in comparisons between ChatGPT and Gemini ($p = 0.154$) or between Gemini and CoPilot ($p = 0.834$). These findings suggest that while CoPilot performs better than ChatGPT, its performance is comparable to that of Gemini.

In the binary classification scenario, the Pearson Chi-square test of independence also indicated a statistically significant relationship between the correct and incorrect responses from ChatGPT and CoPilot ($\chi^2 = 6.545$, $p = 0.011$). Likewise, the comparison results for ChatGPT and Gemini ($\chi^2 = 6.545$, $p = 0.011$) and for CoPilot and Gemini ($\chi^2 = 100.00$, $p < 0.001$) demonstrated significant relationships. All comparisons among the generative AI tools in binary classification illustrate a significant correlation concerning correct and incorrect responses. There is notably high response consistency, especially between CoPilot and Gemini.

Conclusion

The purchasing process is among the most important for a business. Especially in production enterprises, in terms of production planning, regular and balanced purchasing processes should be carried out and monitored. Moreover, artificial intelligence is supposed to be used as a supportive tool in these processes. Thus, this study examines the usability of artificial intelligence tools as decision support in prioritizing the purchase requests generated by users in a manufacturing enterprise. In accordance with this purpose, three large language models: ChatGPT, CoPilot, and Gemini, were tested on 100 requisitions classified by human experts. First, the classification performance of each model was evaluated in the categories of "Urgent," "Normal," and "Not Urgent," and subsequently in "Urgent" and "Normal," using various metrics (accuracy, Cohen's Kappa, precision, recall, F1-score).

The study's findings show that the successful classification percentages of ChatGPT, CoPilot, and Gemini were 53%, 61%, and 61%, respectively. Although CoPilot and Gemini seem to have the same correct classification success, it can be interpreted that they should be the least preferred due to their low precision and recall values. ChatGPT seems to be a cautious model that avoids misclassifications according to metrics. On the other hand, in the binary classification experiments, the highest accuracy was obtained by the CoPilot model (0.83), followed closely by ChatGPT (0.81) and Gemini (0.81). CoPilot seems to be the model with the highest accuracy and consistency, achieving the highest success rate for both classifications. The notably lower performance metrics than the three-class classification can be attributed to data imbalance and class overlap. The scarcity of examples representing some classes limited model learning, and the semantic proximity between the “Normal” and “Not Urgent” categories increased misclassification rates. Moreover, the absence of contextual business data—such as stock levels, production schedules, and procurement lead times—restricted the models’ ability to infer actual priority levels accurately. Nevertheless, all three models demonstrated high performance in binary classification, indicating their robustness when the decision boundary is more precise. On the other hand, the relationships among the generative artificial intelligence tools were statistically evaluated using the Pearson Chi-Square test of independence, at both triple and binary classification levels. The triple classification analysis identified a significant association between ChatGPT and Copilot ($\chi^2 = 9.927$, $p=0.002$). However, no statistically significant relationships were found between ChatGPT and Gemini ($p=0.154$) or between Copilot and Gemini ($p=0.834$). In the binary classification analysis, the association between ChatGPT and Copilot remained statistically significant ($\chi^2 = 6.545$, $p=0.011$); a significant association was also observed between ChatGPT and Gemini ($\chi^2=6.545$, $p=0.011$). The strongest association was detected between Copilot and Gemini ($\chi^2 = 100.000$, $p<0.001$), indicating a high degree of consistency and similarity in their response patterns. These findings demonstrate that the distributions of correct and incorrect responses among generative AI tools are significantly interrelated, with powerful alignment observed between Copilot and Gemini regarding response consistency and decision-making behavior.

Previous studies have shown varying accuracy rates for different AI models. Şensoy and Çıtırık (2025) reported that ChatGPT-3.5 achieved an accuracy rate of 68.8%, CoPilot reached 56.3%, and Gemini scored 43.8%. In contrast, Aydın et al. (2024) found that Gemini had an accuracy of 87.5%, while ChatGPT-3.5 had 45%, ChatGPT-4.0 had 52.5%, and CoPilot achieved 60%. Additionally, Gelmiş et al. (2025) demonstrated that ChatGPT-4 0 outperformed Gemini in their comparison. Overall, these studies indicate that different tools can have varying levels of success, supporting the conclusion that no generative AI model consistently outperforms others across all classification contexts. Furthermore, data quality and task design appear to be the primary factors influencing success.

The study provides an example of the use of generative artificial intelligence in enterprise business processes. For a problem such as the procurement process, this study makes important contributions to the literature and practice. However, the study has some limitations. The sample size is relatively small, and the internal decision mechanisms of the models are not inexplicable. In addition, classifications made without access to contextual business data (inventory status, production plan, lead time, etc.) can sometimes lead to inaccurate results.

When incorporating this work into corporate operations, it is crucial to consider several ethical and security issues. Relying solely on generative AI tools for critical decision-making poses risks related to transparency, accountability, and bias. The "black box" nature of generative AI systems raises questions about who will take responsibility in the event of an unfavorable outcome. Additionally, since these systems learn from past transactions, they are significantly influenced by biased or incomplete datasets.

From a security standpoint, directly integrating generative AI tools with ERP systems requires specific measures to safeguard data privacy and security. Sharing sensitive business information, such as purchasing data, with third-party AI providers can increase the risk of data leaks. Therefore, it is essential to implement necessary precautions, such as data anonymization and encryption.

Moreover, the results generated by AI tools should be verified under human supervision. Preventing operational disruptions caused by misclassifications is vital for maintaining continuous processes. This study suggests that generative AI tools are better suited for decision support when supervised by expert personnel, rather than being used autonomously in purchasing processes. CoPilot, which shows higher classification accuracy, may be preferred in critical production environments, while ChatGPT and Gemini offer different integration models.

Regardless of the tool selected, all necessary precautions must be taken. Future research could replicate similar analyses with larger datasets, utilize specially trained AI models, and explore hybrid decision systems based on human-AI collaboration. Additionally, studies could consider other factors, such as speed, ease of integration, and cost, alongside the accuracy of classification.

AUTHOR STATEMENT

Research and Publication Ethics Statement: This study was prepared according to scientific research and publication ethics rules.

Ethics Committee Approval: This study does not require ethics committee approval, as it does not include analyses that require ethics committee approval.

Author Contributions: The contribution of the author is 100%

Conflict of Interest: The study did not involve conflicts of interest for the author or third parties.

References

Ahmadi, S. (2023). Open AI and its impact on fraud detection in the financial industry. *Journal of Knowledge Learning and Science Technology ISSN: 2959–6386 (Online)*, *2*(3), 263–281.

Alhur, A. (2024). Redefining healthcare with artificial intelligence (AI): The contributions of ChatGPT, Gemini, and Co-Pilot. *Cureus*, *16*(4), e57795. https://doi.org/10.7759/cureus.57795

Al-Worafi, Y., Hermansyah, A., Tan, C., Choo, C., Bouyahya, A., Paneerselvam, G., & Ming, L. (2023). Applications, benefits, and risks of ChatGPT in medical and health sciences research: An experimental study. *Progress in Microbes & Molecular Biology*, *6*(1), 1–13. https://doi.org/10.36877/pmmb.a0000337

Aydın, F. O., Aksoy, B. K., & Ceylan, A. (2024). Refraktif cerrahide sık sorulan sorular için ChatGPT 3.5, ChatGPT 4.0, Gemini ve Microsoft CoPilot tarafından oluşturulan yanıtların okunabilirliği ve uygunluğu. *Turkish Journal of Ophthalmology*, *54*(6), 313-317. https://doi:10.4274/tjo.galenos.2024.28234.

Ayhan, M., & Kılıç, Z. (2024). Yapay zekâ modellerinin hadis tarihi sorularına verdiği yanıtların karşılaştırmalı analizi: ChatGPT ve Gemini örneği. *Dinbilimleri Akademik Araştırma Dergisi*, *24*(3), 137-159.

Banimelhem, O., & Amayreh, W. (2023). *The performance of ChatGPT in emotion classification*. The 14th International Conference on Information and Communication Systems (ICICS), Irbid, Jordan.

Boduroğlu, E., Koç, O., & Yiğiter, M. S. (2023). Madde güçlüklerinin tahmin edilmesinde uzman görüşleri ve chatgpt performansının karşılaştırılması / comparison of expert opinions and ChatGPT performance in predicting item difficulties. *Disiplinlerarası Eğitim Araştırmaları Dergisi*, *7*(15), 202-210. https://doi.org/10.57135/jier.1296255

Bozkurt, A. (2023). Generative artificial intelligence (AI) powered conversational educational agents: The inevitable paradigm shift. *Asian Journal of Distance Education*, *18*(1), 198-204. https://doi.org/10.5281/zenodo.7716416

Brigo, F., Broggi, S., Strigaro, G., Olivo, S., Tommasini, V., Massar, M., ... & Zaboli, A. (2025). Artificial intelligence (ChatGPT 4.0) vs. Human expertise for epileptic seizure and epilepsy diagnosis and classification in Adults: An exploratory study. *Epilepsy & Behavior*, *166*, 110364.

Brynjolfsson, E., Li, D., & Raymond, L. R. (2023). *Generative AI at work (No. w31161)*. National Bureau of Economic Research.

Bulut, C. (2025). Sağlık kurumlarında yapay zekâ destekli karar destek sistemlerinin kullanımı. *Uluslararası Sağlık Yönetimi ve Stratejileri Araştırma Dergisi*, *11*(1), 27-37.

Butean, O., Felek, T., Erkal, D., & Er, K. (2025). *Comparison of responses from different large language models on dens invaginatus cases: a pilot study*. Recep Tayyip Erdoğan Üniversitesi Diş Hekimliği Fakültesi 3. Ulusal Öğrenci Kongresi, Rize, Turkey.

Chatterjee, S., Liu, C., Rowland, G., & Hogarth, T. (2024). *The impact of AI tools on engineering at ANZ Bank: An empirical study on GitHub CoPilot within a corporate environment*. 10th International Conference on Software Engineering (SEC 2024), Melbourne, Australia. https://doi.org/10.5121/csit.2024.140702

Chu, M. N. (2023). Assessing the benefits of ChatGPT for business: an empirical study on organizational performance. *IEEE Access*, *11*, 76427–76436.

Coşar, M. (2024). Kurumsal bilgi güvenliği yönetiminde yapay zekâ destekli risk analizi. *Denetişim*, *31*(1), 144-155. https://doi.org/10.58348/denetisim.1519578

Crewson, P. E. (2005). Reader agreement studies. *American Journal of Roentgenology*, *184*(5), 1391–1397.

Çeber, B. (2024). Reklam ajanslarında yapay zekâ kullanımı: sektör profesyonellerinin ChatGPT ve Midjourney deneyimlerine yönelik bir araştırma. *Erciyes İletişim Dergisi*, *11*(2), 583-606. https://doi.org/10.17680/erciyesiletisim.1439479

Çelik, A., Balcıoğlu, Y., & Altındağ, E. (2025). Understanding user engagement with Gemini through sentiment analysis and concept mapping. *International Journal of Business Analytics*, *12*(1), 1-14. https://doi.org/10.4018/ijban.375429

Dabbas, W. F., Odeibat, Y. M., Alhazaimeh, M., Hiasat, M. Y., Alomari, A. A., Marji, A., ... & Momani, G. M. (2024). Accuracy of ChatGPT in neurolocalization. *Cureus*, *16*(4), e59143. https://doi.org/10.7759/cureus.59143

Dalgıç, A. (2023). Restoran işletmelerine yapılan olumsuz yorumları ChatGPT değerlendirebilir mi? TripAdvisor'da bir uygulama. *İşletme Araştırmaları Dergisi*, *15*(4), 3069-3080.

Darlan, D., Ajani, O. S., An, J. W., Bae, N. Y., Lee, B., Park, T., & Mallipeddi, R. (2025). SmartBerry for AI-based growth stage classification and precision nutrition management in strawberry cultivation. *Scientific Reports, 15*(1), 14019. https://doi.org/10.1038/s41598-025-97168-z

Das, D., Kumar, N., Longjam, L.A., Sinha, R., Deb Roy, A., Mondal, H., et al., (2023). Assessing the capability of ChatGPT in answering first- and second-order knowledge questions on microbiology as per competency-based medical education curriculum. *Cureus,* 15(3), e36034.

Dave, T., Athaluri, S., & Singh, S. (2023). ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations. *Frontiers in Artificial Intelligence, 6*. https://doi.org/10.3389/frai.2023.1169595

Deloitte AI Institute. (2023). *A new frontier in artificial intelligence: Implications of Generative AI for businesses*. Deloitte. https://www.deloitte.com/content/dam/assets-zone3/us/en/docs/services/consulting/2024/us-ai-institute-generative-artificial-intelligence.pdf

Dis, E., Bollen, J., Zuidema, W., Rooij, R., & Bockting, C. (2023). Chatgpt: five priorities for research. *Nature, 614*(7947), 224-226. https://doi.org/10.1038/d41586-023-00288-7

Duarte, F. (2025). *Number of ChatGPT users (July 2025)*. Exploding Topics. https://explodingtopics.com/blog/chatgpt-users

Emekli, E., Özel, B., & Emekli, E., (2025). *Psikiyatri eğitiminde klinik akıl yürütme becerilerinin geliştirilmesi: chatgpt ile otomatik senaryo ve soru üretimi*. 10. Psikiyatri Zirvesi ve 17. Anksiyete Kongresi, Antalya, Turkey.

En, L., Tha, N., & Thuan, T. (2024). Exploring the application of ChatGPT in learning English of students – a study on perception of sophomores at Nguyen Tat Thanh University. *Journal of Science and Technology, 7*(3), 41-47. https://doi.org/10.55401/fg15by35

Erdem, E. (2024a). Yapay zeka uygulamalarinin sosyal bilim alaninda yapilan çalişmalarda uygulanabilirliği: Chatgpt, Bing ve Youchat örneği. *İletişim Bilimi Araştırmaları Dergisi, 4*(3), 218-234.

Erdem, M. B. (2024b). Lojistik firma web sitelerinin chatgpt ile analiz edilmesi. *Erciyes Akademi, 38*(4), 1085-1099. https://doi.org/10.48070/erciyesakademi.1596551

Fabijan, A., Polis, B., Fabijan, R., Zakrzewski, K., Nowosławska, E., & Zawadzka-Fabijan, A. (2023). Artificial intelligence in scoliosis classification: an investigation of language-based models. *Journal of Personalized Medicine, 13*(12), 1695. https://doi.org/10.3390/jpm13121695

Gelmiş, M., Ayten, A., Özsoy, Ç., Bulut, B., & Köse, M. G. (2025). Comparing ChatGPT and Google Gemini in urology: Which AI model provides superior patient education on penile prosthesis?. *Androloji Bülteni, 27*(1), 8-12.

Göktaş, L. S. (2023). ChatGPT uzaktan eğitim sınavlarında başarılı olabilir mi? Turizm alanında doğruluk ve doğrulama üzerine bir araştırma. *Journal of Tourism and Gastronomy Studies, 11*(2), 892-905. https://doi.org/10.21325/jotags.2023.1224

Gupta, R., Hamid, A., Jhaveri, M., Patel, N., & Suthar, P. (2024). Comparative evaluation of AI models such as ChatGPT 3.5, ChatGPT 4.0, and Google Gemini in neuroradiology diagnostics. *Cureus, 16*(8): e67766 https://doi.org/10.7759/cureus.67766

Güney, A., & Özcan, E. (2025). Yapay zekanın muhasebede kullanımı üzerine bir araştırma. *Journal of Accounting Institute, (72),* 1-13. https://doi.org/10.26650/MED.1492095

Hacıbey, İ. & Halis, A. (2025). Assessment of artificial intelligence performance in answering questions on onabotulinum toxin and sacral neuromodulation. *Investigative and Clinical Urology, 66*(3), 188-193. https://doi.org/10.4111/icu.20250040

Harada, T., Sato, S. & Nishiura, M. (2025). Using generative AI to ımprove library book classification accuracy: the role of ıncreased training data. Oliver, G., Frings-Hessami, V., Du, J.T., Tezuka, T. (eds.), *Sustainability and empowerment in the context of digital libraries*. ICADL 2024. Lecture Notes in Computer Science, vol 15494. Springer, Singapore. https://doi.org/10.1007/978-981-96-0868-3_20

Huang, S., Grady, P., & GPT-3. (2022, September 19). *Generative AI: a creative new world*. Sequoia. https://www.sequoiacap.com/article/generative-ai-a-creative-new-world/

Hussain, A., & Ahmed, A. (2023). *Auto-classification of harmonized tariff codes using ChatGPT*. 4th International Conference on Distributed Sensing and Intelligent Systems (ICDSIS 2023), Dubai, UAE.

İbiş, S. (2025). Yapay zekâ teknolojilerinin gastronomi turizminde kullanımı: ChatGPT örneği. *Güncel Turizm Araştırmaları Dergisi, 9*(1), 109-131. https://doi.org/10.32572/guntad.1518594

Jalil, S., Rafi, S., LaToza, T. D., Moran, K., & Lam, W. (2023). *ChatGPT and software testing education: Promises & perils*. IEEE International Conference on Software Testing, Verification, and Validation Workshops (ICSTW), Dublin, Ireland.

Kaftan, A., Hussain, M., & Naser, F. (2024). Response accuracy of ChatGPT 3.5 CoPilot and Gemini in interpreting biochemical laboratory data: A pilot study. *Scientific Reports, 14*(1), https://doi.org/10.1038/s41598-024-58964-1

Kaiser, K., Hughes, A., Yang, A., Turk, A., Mohanty, S., Gonzalez, A., … & Ellis, R. (2024). Accuracy and consistency of publicly available large language models as clinical decision support tools for the management of colon cancer. *Journal of Surgical Oncology, 130*(5), 1104–1110. https://doi.org/10.1002/jso.27821

Kaplan, A., & Haenlein, M. (2019). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons, 62*(1), 15–25.

Koubaa, A. (2023). GPT-4 vs. GPT-3.5: A concise showdown. *Prerprints*. https://doi.org/10.20944/preprints202303.0422.v1

Kutlucan, E., & Seferoğlu, S. S. (2024). Eğitimde yapay zekâ kullanımı: ChatGPT'nin KEFE ve PEST analizi. *Türk Eğitim Bilimleri Dergisi, 22*(2), 1059-1083. https://doi.org/10.37217/tebd.1368821

Küçük, D., & Fazlı, C. (2024). Hukuki metinlerin otomatik işlenmesinde yapay zekâ teknolojilerinin kullanımı. *Bilişim Hukuku Dergisi, 6*(1), 1-23. https://doi.org/10.55009/bilisimhukukudergisi.1450588.

Küçüker, M. (2023). Muhasebede yapay zekâ uygulamaları: ChatGPT'nin muhasebe sınavı. *Fırat Üniversitesi Sosyal Bilimler Dergisi, 33*(2), 875-888. https://doi.org/10.18069/firatsbed.1289885

Lawton, G. (2023, July 20). *What is generative AI? Everything you*. Techtarget. https://www.techtarget.com/searchenterpriseai/definition/generative-AI

Lee, K., Kessler, D., Caglič, I., Kuo, Y., Shaida, N., & Barrett, T. (2024). Assessing the performance of ChatGPT and Bard/Gemini against radiologists for prostate imaging-reporting and data system classification based on prostate multiparametric MRI text reports. *British Journal of Radiology, 98*(1167), 368–374. https://doi.org/10.1093/bjr/tqae236

Li, M., & Guenier, A. (2024). ChatGPT and health communication. *International Journal of E-Health and Medical Communications, 15*(1), 1–26. https://doi.org/10.4018/ijehmc.349980

Liu, C., Wei, M., Qin, Y., Zhang, M., Jiang, H., Xu, J., ... & Zhou, J. (2024). Harnessing large language models for structured reporting in breast ultrasound: a comparative study of OpenAI (GPT-4.0) and Microsoft Bing (GPT-4). *Ultrasound in Medicine & Biology, 50*(11), 1697-1703.

Liu, Y., He, S., Chen, H., Qing, W., Su, H., & Deng, G. (2023). Investigation on response strategies for the impact of ChatGPT technology application on college Chinese language education. *International Journal of Educational Innovation and Science, 4*(1), 156–164. https://doi.org/10.38007/ijeis.2023.040114

Makridakis, S. (2017). Forecasting the impact of artificial intelligence (AI). *Foresight: The International Journal of Applied Forecasting*, (47), 7-13.

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (1955). A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955. *AI Magazine, 27*(4), 12.

Mich, L., & Garigliano, R. (2023). ChatGPT for e-tourism: A technological perspective. *Information Technology & Tourism, 25*, 1-12. https://doi.org/10.1007/s40558-023-00248-x

Moor, J. (2006). The Dartmouth College artificial intelligence conference: The next fifty years. *AI Magazine, 27*(4), 87–87.

Murindanyi, S., Wycliff Mugalu, B., Nakatumba-Nabende, J., & Marvin, G. (2023). *Interpretable machine learning for predicting customer churn in retail banking. 7th International Conference on Trends in*

Electronics and Informatics (ICOEI), Tirunelveli, India. https://doi.org/10.1109/ICOEI56765.2023.10125859

Okur, H., & Ekşi, A. (2024). Yapay zekâ (AI) teknolojilerinin islam eşya hukuku bilgisi üzerine bir değerlendirme: Chatgpt ve Google Gemini karşılaştırması. *Dinbilimleri Akademik Araştırma Dergisi*, *24*(3), 29-54. https://doi.org/10.33415/daad.1582059.

Osama, M., & Afridi, S. (2023). ChatGPT: a new era in research writing assistance. *Journal of the Pakistan Medical Association, 73*(9), 1929-1930. https://doi.org/10.47391/jpma.9183

Pavlik, V. J. (2023). Collaborating with ChatGPT: considering the implications of generative artificial intelligence for journalism and media education. *Journalism & Mass Communication Educator*, *78*(1), 84-93.

Precedence Research. (2025, May 22). Generative *aı market. Generative aı market size to hit USD 1005.07 bn by 2034*. Precedence Research. https://www.precedenceresearch.com/generative-ai-market

Rane, N., Choudhary, S., & Rane, J. (2024). Gemini or ChatGPT? Capability, Performance, and Selection of Cutting-edge Generative Artificial Intelligence (AI) in Business Management. *Studies in Economics and Business Relations*, *5*(1), 40–50. https://doi.org/10.48185/sebr.v5i1.1051

Rossettini, G., Rodeghiero, L., Corradi, F., Cook, C., Pillastrini, P., Turolla, A., … & Palese, A. (2024). Comparative accuracy of ChatGPT-4, Microsoft CoPilot, and Google Gemini in the Italian entrance test for healthcare sciences degrees: a cross-sectional study. *BMC Medical Education*, *24*(1), 1-9 https://doi.org/10.1186/s12909-024-05630-9

Sabir, A., Ali, H. A., & Aljabery, M. A. (2024). ChatGPT tweets sentiment analysis using machine learning and data classification. *Informatica*, *48*(7), 103–112. https://doi.org/10.31449/inf.v48i7.5535

Sallam, M. (2023). ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. *Healthcare, 11*(6), 887. https://doi.org/10.3390/healthcare11060887

Salman, I. M., Ameer, O. Z., Khanfar, M. A., & Hsieh, Y-H. (2025). Artificial intelligence in healthcare education: evaluating the accuracy of ChatGPT, CoPilot, and Google Gemini in cardiovascular pharmacology. *Frontiers in Medicine, 12,* 1495378. https://doi: 10.3389/fmed.2025.1495378

Shaheen, A., Afflitto, G. G., & Swaminathan, S. S. (2025). ChatGPT-Assisted classification of postoperative bleeding following microinvasive glaucoma surgery using electronic health record data. *Ophthalmology Science, 5*(1), 100602.

Shihab, R., Sultana, N., & Samad, A. (2023). Revisiting the use of ChatGPT in business and educational fields: possibilities and challenges. *BULLET: Jurnal Multidisiplin Ilmu*, *2*(3), 534–545. https://journal.mediapublikasi.id/index.php/bullet/article/view/2761

Smales, L. A. (2023). Classification of RBA monetary policy announcements using ChatGPT. *Finance Research Letters, 58*, 104514. https://doi.org/10.1016/j.frl.2023.104514

Sudirjo, F., Diantoro, K., Al-Gasawneh, J. A., Khootimah Azzaakiyyah, H., & Almaududi Ausat, A. M. (2023). Application of ChatGPT in ımproving customer sentiment analysis for businesses. *Jurnal Teknologi Dan Sistem Informasi Bisnis*, *5*(3), 283-288. https://doi.org/10.47233/jteksis.v5i3.871

Şensoy, E., & Çıtırık, M. (2025). Glokom konusunda yapay zeka sohbet botlarının performans analizi: ChatGPT-3.5, Gemini ve CoPilot'un İngilizce ve Türkçe sorulardaki başarı farklılıkları. *MN Oftalmoloji*, *32*(1), 17-21.

Thormnundsson, B. (2023, June 20). *ChatGPT-statistics & facts*. Statista. https://www.statista.com/topics/10446/chatgpt/#topicOverview

Uyar, T. (2024). Chat gpt'nin serbest mantıksal safsata tespitinde kullanımı. *Yeni Medya Elektronik Dergisi*, *8*(1), 144-179.

Ventura, M., & Filho, A. (2023). ChatGPT: limitations, challenges, and potential applications. *Brazilian Journal of Science*, *3*(1), 65–68. https://doi.org/10.14295/bjs.v3i1.427

Waisberg, E., Ong, J., Masalkhi, M., Zaman, N., Sarker, P., Lee, A. G., & Tavakkoli, A. (2024). Google's AI chatbot "Bard": a side-by-side comparison with ChatGPT and its utilization in ophthalmology. *Eye*, *38*(4), 642–645. https://doi.org/10.1038/s41433-023-02760-0

Wiles, J. (2023, January 26). *Beyond Chatgpt: the future of generative AI for enterprises*. Gartner. https://www.gartner.com/en/articles/beyond-chatgpt-the-future-of-generative-ai-for-enterprises

Wong, M., Guo, S., Hang, C., Ho, S., & Tan, C. (2023). Natural language generation and understanding big code for AI-assisted programming: a review. *Entropy, 25*(6), 888. https://doi.org/10.3390/e25060888

Wulcan, J. M., Jacques, K. L., Lee, M. A., Kovacs, S. L., Dausend, N., Prince, L. E., ... & Keller, S. M. (2025). Classification performance and reproducibility of GPT-4 Omni for information extraction from veterinary electronic health records. *Frontiers in Veterinary Science, 11*, 1490030.

Guo, Y., & Wan, Z. (2024). *Performance Evaluation of Multimodal Large Language Models (LLaVA and GPT-4-based ChatGPT) in Medical Image Classification Tasks*. 12th International Conference on Healthcare Informatics (ICHI), Orlando, FL, USA. Doi: 10.1109/ICHI61247.2024.00080.

Dökme Yağar, S. (2023). ChatGPT'nin sağlık alanındaki potansiyel kullanımına ilişkin çıkarımlar. *Business & Management Studies: An International Journal, 11*(3), 1226–1240. https://doi.org/10.15295/bmij.v11i3.2264.

Yalçın-Çelik, A., & Çoban, Ö. K. (2023). Kimya sorularının cevaplanmasında yapay zekâ tabanlı sohbet robotlarının performansının incelenmesi. *Journal of Turkish Educational Sciences, 21*(3), 1540-1561.

Yılmaz, M., & Çil, G. (2024). Yapay zekâ, erektil disfonksiyon ile ilgili sorulara nasıl yanıtlar veriyor: ChatGPT vs Google AI. *Androloji Bülteni, 26*(2), 116−122. https://doi.org/10.24898/tandro.2024.12989

Zhang, B., Liang, P., Zhou, X., Ahmad, A., & Waseem, M. (2023). Demystifying practices, challenges, and expected features of using GitHub CoPilot. *International Journal of Software Engineering and Knowledge Engineering, 33*(11n12), 1653–1672. https://doi.org/10.1142/s0218194023410048

Zhu, J. J., Jiang, J., Yang, M., & Ren, Z. J. (2023). ChatGPT and environmental research. *Environmental Science & Technology, 57*. 1-4. https://doi.org/10.1021/acs.est.3c01818