AI for Knowledge-Driven Business Decisions: Capabilities and Constraints

Giovanni De Pretto[1], Ettore Bolisani[1] and Maayan Nakash[1,2]
[1]Department of Management and Engineering, University of Padova, Vicenza, Italy
[2]Department of Management, Bar-Ilan University, Israel

Giovanni.Depretto@studenti.unipd.it
Ettore.Bolisani@unipd.it
Maayan.Nakash@biu.ac.il

Abstract: This paper presents experimental research evaluating the capability of generative artificial intelligence (GenAI) chatbots to assist business executives in their decision-making processes. With the growing prominence of large language models such as ChatGPT and Perplexity, there is significant interest in their potential for knowledge management (KM) and decision support, but empirical evidence is still lacking. This study conducted over 210 experiments using approximately 900 prompts to assess the performance of ChatGPT and Perplexity across various business decision tasks, including marketing analysis, work shift optimization, sentiment analysis, review of company financial and strategic documentation, and product marketing evaluation. The outputs generated by the chatbots were compared with traditional decision support methods, human expert analysis, and benchmarking data. The results indicate that GenAI chatbots can streamline processes by collecting, interpreting, and synthesizing data into actionable insights. Nevertheless, limitations were identified in handling complex file formats, response variability based on context framing, and the necessity for human validation of outputs. A comparative analysis of ChatGPT and Perplexity revealed distinct differences: ChatGPT demonstrated overall reliability but struggled with tabular data, whereas Perplexity provided more substantive yet occasionally less accurate responses. The findings suggest that GenAI chatbots could enhance decision efficiency through rapid analysis, but a collaborative human-AI model remains advisable to validate outputs and iteratively refine queries. Our study underscores the potential of GenAI for KM in guiding decisions while highlighting the need for further research on evaluation methods, integration with business intelligence tools, and the development of user guidelines for the responsible adoption of this technology.

Keywords: Applied knowledge management, Business decision making, Chatbots, Experiments, GenAI

1. Introduction

In recent years, artificial intelligence (AI) powered by large language models (LLMs) has gained considerable traction in both public and organizational contexts. Intuitive interfaces, natural language interaction capabilities, and cost-efficiency have positioned it as a promising technology for knowledge management (KM) applications (Alavi et al. 2024; Jarrahi et al. 2023; Kirchner et al. 2024; Nakash and Bolisani, 2024; Storey, 2025). These systems are increasingly viewed as potential decision-support instruments for business managers, owing to features such as multidimensional data analysis, interactive prompting, complex question answering, and the ability to synthesize new knowledge through the recombination of existing information (Shrestha et al. 2019). However, despite their growing visibility and adoption, there remains limited clarity regarding the actual efficacy of such systems in handling complex, domain-specific knowledge and facilitating high-quality managerial decision-making.

While current literature acknowledges the theoretical potential of generative AI (GenAI) for augmenting knowledge-based decisions, much of the discourse remains conceptual, technical, or anecdotal in nature (Duan et al. 2019). Empirical studies assessing the performance of GenAI systems under realistic conditions, particularly in business decision-making scenarios, are scarce. Moreover, the capabilities of these tools can vary widely depending on task complexity, domain specificity, and quality of training data (Stray, 2021). These uncertainties are further compounded by concerns about reliability, transparency, and algorithmic bias, which may undermine trust in AI-generated outputs (Arrieta et al. 2020; Gichoya et al. 2023).

This study addresses these gaps and aims to empirically assess whether and how generative AI chatbots can support business decision-making processes, particularly in knowledge-intensive contexts. While generative models have been praised for their potential in knowledge management, there is still a lack of systematic experimental studies evaluating their performance under real-world conditions. Specifically, the following research questions guided the investigation:

RQ1: How effective are GenAI chatbots in executing business-relevant analytical tasks?

RQ2: How do chatbot-generated results compare to those produced by traditional decision-support methods and human experts?

RQ3: What are the specific strengths and limitations of different GenAI systems (e.g., ChatGPT and Perplexity) in real decision scenarios?

RQ4: What methodological considerations are necessary to ensure a reliable and responsible use of GenAI tools in business environments?

These questions were addressed through a series of experiments simulating typical business situations, as detailed in the following sections. By comparing GenAI outputs to traditional knowledge processing approaches, the research provides a grounded assessment of current AI capabilities and outlines critical issues for their responsible integration in managerial contexts. The study further explores the critical role of human expertise in ensuring the effective utilization of GenAI for decision-making, which is one of the major contemporary challenges in its application to business and management (Kolbjørnsrud et al. 2016).

2. Revolutionizing Business with GenAI

Recent advancements in GenAI have enabled these systems to perform routine tasks such as data classification and reorganization. Their capabilities extend to creating text, composing music, and generating digital art, which has attracted significant attention and prompted independent experimentation by consumers and families. This widespread interest has led stakeholders to consider the impact of GenAI on business and society, often without a comprehensive understanding of its implications (Cano-Marin, 2024; Feuerriegel et al. 2024; Kanbach et al. 2024). Although GenAI technologies still represent a limited portion of overall AI investments, funding for these innovations is growing rapidly. In the first five months of 2023 alone, investments reached $12 billion, with an average annual growth rate of 74% between 2017 and 2022, compared to 29% for AI overall (McKinsey and Company, 2023). The potential revenue from GenAI could reach approximately $1.1 trillion by 2028, up from $45 billion in 2024 (Morgan Stanley Research, 2025).

GenAI has garnered global attention for its potential to transform knowledge-based work across various sectors and business functions (Agrawal et al., 2022). Areas such as sales, marketing, customer operations, and software development could be revolutionized through performance improvements and role optimization. Among the numerous business functions impacted, some stand out for the value generated relative to functional costs. According to McKinsey's analysis, four business functions could represent approximately 75% of the total annual value from GenAI use cases: customer operations, marketing and sales, software engineering, and research and development (R&D). Benefits include automating customer service, creating personalized content and advertisements, accelerating coding and documentation, and generating innovative research hypotheses and designs (McKinsey and Company, 2023).

Beyond functional value, GenAI has the potential to transform organizational KM systems. Its natural language capabilities enable conversational querying to retrieve archived information, allowing employees rapid access to relevant knowledge for informed decisions and strategies (Kolbjørnsrud et al., 2016). In 2012, knowledge workers spent 20% of their time (one workday per week) searching for and gathering information (McKinsey Global Institute, 2012) and according to empirical evidence collected during COVID-19, some workers now spend up to one and a half working days per week on these tasks (Nakash and Bouhnik, 2024).

If GenAI could assume this task, it could significantly boost efficiency and productivity. Moreover, as a "virtual expert," GenAI can analyze vast company data in natural language and interact iteratively with users to refine and personalize searches, offering scalability far beyond dedicated expert teams (Shrestha et al., 2019). Operational benefits include synthesizing reports and data for decision support, answering operational questions without manual database searches, retrieving historical project knowledge, and assisting with consistent, accurate internal document drafting.

These examples underscore GenAI's potential to revolutionize corporate KM by reducing information retrieval time and costs while enhancing operational efficiency organization-wide. However, despite rising interest, there is insufficient empirical evidence regarding the real effectiveness of these tools in supporting managerial decision-making. Recent studies (Cillo and Rubera, 2024; Csaszar et al. 2024; Soni, 2023) have begun analyzing the potential and impact of GenAI on business decisions, but there is a consensus that specific experiments must be conducted to understand the true effectiveness of these technologies, particularly from the perspective of potential users in real-life situations.

3. Research Outline

This study outlines a series of experiments designed to simulate real-world business decision scenarios, reflecting typical business situations. The methodology proceeded as follows.

A: Selection of Realistic Experiments

A series of experiments were selected based on their relevance to business management and potential applications of GenAI. Some scenarios involve quantitative analysis, others require information categorization, and some entail qualitative analysis:

- Simulation of Marketing Models
- Optimization of Work Shifts
- Sentiment Analysis
- Analysis of Company Financial and Strategic Documentation
- Marketing Evaluation of a Product Family

B: Identification of Generative Chatbots

Selecting appropriate generative chatbots posed several challenges. First, technologies evolve fast, and new versions of chatbots are being continuously introduced, with new algorithms and analytic tools. Second, standard criteria to assess and compare chatbot performances are not yet available. Therefore, it was decided to adopt a mixed approach including quantitative and subjective assessments, and also indications from professional experts. These criteria included: the GPQA Diamond applied to chatbots (applied, for instance, in framework like BenchmarkAggregator); the Massive Multitask Language Understanding (MMLU) benchmark; MATH, a MATH comprehensive evaluation tool to assess mathematical reasoning of GenAI) models; Multilingual Grade School Math (MGSM) benchmark; the LMArena score developed by Chatbot Arena; the Olympic Arena benchmark measurement. These measurement methods are still experimental, and each provides a specific perspective. Therefore, an overall view of all these measurements was taken. A third issue is that we had some specific restrictions regarding the access to certain chatbots especially in their “premium” versions, which limited the possible choice. In the end, it was decided to use these tools and versions: ChatGPT Plus and Perplexity Pro that, at least at the time of our study (December 2024) showed the best combination of scores.

C: Selection of Assessment Approaches

Given the lack of standard methods to evaluate quality and effectiveness of GenAI systems in decision-making contexts, two approaches were adopted. One approach involved comparing chatbot results with those obtained using existing decision-making analytical tools. For instance, in marketing-related experiments, input data and benchmarking results were derived from Enginius, a software that performs marketing analysis with predefined models. Thus, chatbot results were compared to those a manager might obtain using traditional decision support systems. In other cases, subjective analysis was performed, such as comparing chatbot analyses of financial and strategic documentation with manual analyses or professional reports of the same documents.

D: Conducting Experiments

For each experiment, a series of prompts were compiled for the two chatbots, and data files were input in various formats.

4. Experiments and Results

4.1 Marketing Analysis

This section details an experiment aimed at comparing the effectiveness of GenAI chatbots with a specialized marketing platform called Enginius. The objective is to evaluate how these solutions interpret marketing data and transform it into useful and accurate outputs. Three models from the Enginius platform were selected: the GE-McKinsey Matrix, Segmentation Analysis, and Network Analysis. For each model, a structured approach was followed, providing key information and an Excel file containing the initial data necessary to address the problem. The process was replicated using Word and PDF files to verify the models' ability to handle different document formats. The analysis of the GE-McKinsey Matrix involved calculating and representing the x and y coordinates for each product family within the matrix, considering evaluation data, weight perspectives, and market shares. ChatGPT achieved 100% accuracy on PDF and Word formats in both languages when calculating

x and y, correctly recognizing scores and applying normalization. Perplexity showed mixed performance, initially struggling with data interpretation but improving with advanced search modes on Excel files.

For Segmentation Analysis, the objective was to identify the optimal number of segments using hierarchical clustering criteria and assign respondents to their respective clusters. Both ChatGPT and Perplexity performed well in determining the number of segments and assigning respondents, but faced challenges in synthesizing data from different sources, with ChatGPT's advanced model (o1) being the only one to achieve maximum accuracy.

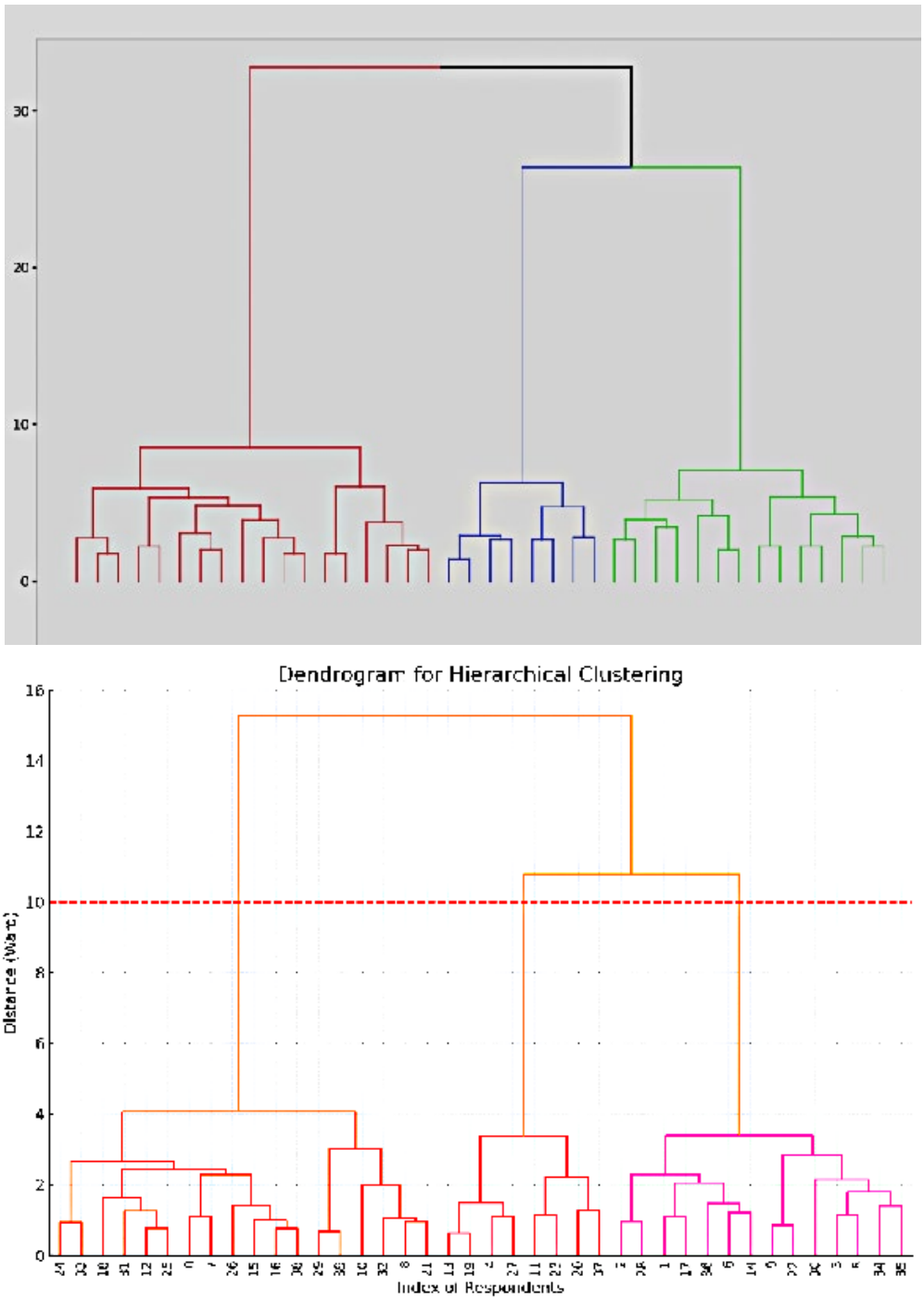

Figure 1: Comparison between dendrogram produced by Enginius (top) and Chatbot (bottom)

A task assigned to AI was the calculation of the optimal number of segments, through the graphical representation of a dendrogram. This visualization provides the business analyst with an immediate visual aid to identify the most appropriate number of segments to adopt (Figure 1). The Network Analysis involved calculating metrics such as degree and closeness centrality for each node in a network of 152 nodes. ChatGPT

excelled with Excel files, accurately calculating metrics and providing graphical representations and statistical summaries. However, it struggled with PDF and Word formats due to difficulties in extracting and interpreting the data matrix correctly. Perplexity was unable to calculate network metrics or provide satisfactory graphical representations in any case.

Overall, the experiments highlight the potential of GenAI chatbots in supporting business decision-making processes, particularly in the initial analysis phases. However, they also emphasize the need for rigorous methodological control, especially when dealing with complex data integration or probabilistic estimates.

4.2 Optimization of Work Shifts

This section describes an experiment aimed at evaluating the effectiveness of GenAI chatbots in optimizing the scheduling of operators in an inbound customer service contact center. The contact center, assumed to manage a company in the water sector, operates with unique toll-free numbers. The experiment utilized an Excel file containing historical work schedules from previous weeks, including operators' experience levels (Coordinator, Lead, Senior, Junior).

The GenAI systems were tasked with predicting and assigning shifts for the "current week" based on the provided information and adhering to precise operational rules, such as shift durations, preferred and less preferred time slots, evening shift rotation, coordinator and lead operator schedules, fair rotation among operators, operating hours, continuous coverage, and call flow peaks. The results were evaluated using both subjective and objective criteria, leveraging the personal experience of one of the paper's authors as a shift scheduler in a similar contact center. The evaluation considered adherence to operational rules, optimization of hourly coverage, fairness of rotations, and overall applicability and reliability. Essentially, the chatbots' results were benchmarked against those that a human manager would have obtained through personal analysis.

The analysis highlights the challenges AI models face in simultaneously applying complex rules when data contains numerous variables. While ChatGPT and Perplexity demonstrated good graphical representation capabilities, they struggled to adhere strictly to stringent constraints, such as preferred time slots and fair rotation. The experiment underscores the need for a "human-assisted" approach in such scenarios, where chatbots can provide baseline solutions or useful visualizations, but human intervention remains crucial for fully reliable results. In other words, human managers can use the systems as "assistants" but must then evaluate their results and, by interacting with the chatbots using different prompts, correct mistakes and inconsistencies to ultimately achieve satisfactory outcomes.

4.3 Sentiment Analysis

This section describes an experiment aimed at examining the behavior of GenAI systems when subjected to a large dataset. Specifically, the study focuses on the analysis of 4,225 comments extracted from Instagram posts about the unveiling of the Ferrari F80, a new supercar revealed on October 17, 2024. Despite the enthusiasm of experts and enthusiasts, the model received several criticisms, mainly regarding three aspects: the controversial design, the transition from a V12 or V8 engine to a V6 hybrid, and the excessive price tag of four million dollars.

The comments were pre-processed by removing those containing only emojis and translating them into English to ensure more accurate analysis. These were provided to the GenAI systems in three different formats (Excel, Word, and PDF) to verify any differences in results based on the input type. The systems were tasked with classifying the comments based on tone (positive, negative, or neutral) and generating a summary table. Additionally, they were asked to create a word cloud to identify the most frequent words and conduct a valence analysis to identify the most commonly used positive, negative, or neutral terms. This approach aimed to provide an overview of users' perceptions and reactions toward the Ferrari F80 for an initial qualitative comparison with the previously described criticisms.

The results of this study were compared with two types of analyses conducted using the Enginius platform: a sentiment analysis on a customized set of comments and an overall sentiment analysis on Instagram using the #ferrarif80 hashtag. Furthermore, a separate evaluation method was developed to directly compare the classification proposed by the chatbots with a human evaluation of a randomly selected sample of 99 comments. These comments were manually analyzed, classifying both the tone (positive, neutral, or negative) and the expressed emotions (anger, anticipation, disgust, fear, joy, and sadness). Figure 2 presents an example of tone and emotion analysis, comparing the results of a manually conducted and an AI-generated assessment.

The evaluation of the results revealed strengths and weaknesses of both ChatGPT and Perplexity. ChatGPT, in the GPT-4o version, excelled at graphical representation, creating clear and intuitive visualizations, sometimes even better than those offered by specialized software. However, its actual analysis capabilities presented many shortcomings, with often unreliable results unless using advanced models like o1 or o3-mini high. Perplexity, on the other hand, showed greater solidity in the overall analysis, providing more coherent results compared to the reference data. However, as the volume of data to be examined increased, the model tended to lose precision, with some gaps in interpreting more complex information.

ID	COMMENT	TONE ANALYSIS	EMOTION ANALYSIS
9	Have you tried the car? No no one has and I fully know this car will be hella fast because of v6 hybrid...	POSITIVE	TRUST
51	Awesome	POSITIVE	JOY
62	Beauty 💀 💀 💀 💀 💀	POSITIVE	JOY
77	Downgrade	NEGATIVE	DISGUST (ANGER)
94	It looks weird but I'll never be able to buy one anyway.	NEUTRAL	SADNESS (ANTICIPATION)
99	V6 hybrid? Well, at least is powerful and beautiful.	NEUTRAL	ANTICIPATION (SURPRISE or JOY)
113	When LaFerrari was launched, it was criticised for design...	NEUTRAL	TRUST
122	Hey Ferrari...why you hiding your tiny engine at the back? 😂 no glass cover?	NEGATIVE	ANGER (DISGUST)
129	idk, it honestly looks dope	POSITIVE	JOY

Comment ID	Comment	Sentiment	Emotion
9	Have you tried the car? No no one has and i fully know this car will be hella fast because of v6 hybrid...	Positive	Trust
51	Awesome	Positive	Joy
62	Beauty 😜 😜 😜 😜 😜	Positive	Joy
77	Downgrade	Negative	Disgust
94	It looks weird but I'll never be able to buy one anyway.	Neutral	Sadness
99	V6 hybrid? Well, at least is powerfull and beautiful.	Neutral	Surprise
113	When laferrari was launched, it was criticised for design...	Neutral	Trust
122	Hey Ferrari...why you hiding your tiny engine at the back? 😂 no glass cover?	Negative	Disgust
129	idk, it honestly looks dope	Positive	Joy

Figure 2: Comparison of tone and emotion analyses (manual assessment – top; chatbot-generated – bottom)

Additional attempts were made to improve the accuracy of the analyses provided by the chatbots, including the use of a reference database and a detailed description of the factors negatively influencing people's perceptions. While these efforts yielded some improvements, particularly for Perplexity, the overall performance remained limited, highlighting the need for further advancements in GenAI models to handle complex sentiment analysis tasks effectively.

4.4 Analysis of a Company's Financial and Strategic Documentation

The fourth experiment aimed to explore the potential of GenAI tools in analyzing a diverse set of heterogeneous and not necessarily directly related documents. Specifically, the investigation focused on the documentation of Piaggio & C., a historic Italian manufacturer of two and three-wheeled motor vehicles. The analyzed material was collected from the company's official website and includes press releases, conference call transcripts and presentations, codes of ethics, interim management reports, corporate governance reports, remuneration policies, financial statements, and non-financial declarations.

The analysis aimed to deepen the understanding of how AI technologies can interpret and leverage complex and varied information. The chatbots were tasked with assuming the role of expert financial analysts to assess the company's health. The analyses focused on financial ratios, key points from shareholder meetings, non-financial performance indicators and ESG policies, identification of risks and opportunities, future strategies and objectives, and an overall assessment of the company's health with a focus on potential future crises.

The accuracy of GenAI chatbots was evaluated using a combination of quantitative and qualitative methods, including comparison with historical data from specialized databases and direct review made by expert analysts. The evaluation of the results revealed both strengths and weaknesses of the two chatbots, ChatGPT and Perplexity. ChatGPT excelled in calculation accuracy and the development of extended and structured analyses, also offering targeted operational suggestions for improving internal management and corporate strategy. Perplexity, while adopting a more concise and sometimes less rigorous approach in the numerical part, proved equally reliable in identifying key themes and contributed additional insights.

Overall, both models provided responses in line with the expectations of human analysis, highlighting risks and opportunities (such as volatility in global markets, electrification, and premium branding strategies), as well as major warning signs (e.g., a possible slowdown in sales in some geographic areas or high levels of debt). This convergence of results suggests that chatbots can complement and, in part, accelerate the work of analysts and consultants, allowing them to focus on strategic insights rather than the operations of collecting and synthesizing information. However, a problem arises from the lack of a standard way to evaluate the assessments provided by the system, and without cross-checking by human decision-makers, there may be a risk of blindly trusting the system.

Looking ahead, it is conceivable that GenAI tools will evolve further, improving numerical precision and refining the ability to integrate information from multiple documentary sources. This could translate into greater effectiveness in identifying emerging trends, anticipating risks, and proposing operational solutions for companies, making AI a key element in the decision-making process.

4.5 Marketing Evaluation of a Product Family

The primary objective of this final analysis is to understand how GenAI chatbots respond to the strategic analysis of a product family. Specifically, the investigation focuses on comparing technical specifications, informative documents, and web pages of a specific line of devices for cyclists—the Garmin Edge line, an advanced technological product line developed by Garmin, a leading company in the fitness and outdoor technology sector. The analysis was conducted in both Italian and English to evaluate the differences in the response capabilities of AI systems based on language. The products under investigation are Garmin Edge 540, 840, and 1050, all consumer products (GPS cycling computers).

This type of analysis is strategic and innovative in nature for decision-making. It enables marketing managers to understand how the products differ from each other and from the competition, highlighting distinctive features that could represent a competitive advantage. The objective of this marketing analysis is also to identify strategies to improve market competitiveness through proposals for technological innovation and targeted positioning. Additionally, the analysis explores development opportunities for each model, considering emerging trends in the relevant sector and the specific needs of different customer segments. Finally, it aims to identify the most effective strategies to differentiate from the competition in terms of innovation, pricing, and related services to strengthen the company's market position.

The chatbots were tasked with developing:

- A SWOT analysis for each model, identifying strengths, weaknesses, opportunities, and threats.
- A comparison with main competitors, identifying direct and indirect competitors and analyzing their offerings in terms of technology, price, features, and services.

- Industry benchmarking, comparing the technical specifications of Garmin models with market leaders to highlight competitive advantages.
- Technological advantages, exploring the technologies integrated into each model compared to alternatives on the market.
- Innovation proposals, suggesting technological improvements for each model.
- Risk analysis and mitigation strategies, identifying potential technological, market, and regulatory risks that could impact sales or reputation.

To verify the accuracy of the results, it is essential to adopt a structured approach that considers various critical aspects, such as completeness, relevance, and quality of the responses. The evaluation of the results shows that both ChatGPT and Perplexity are capable of successfully carrying out the analysis of a product family, starting from a detailed description, as in this case through user manuals. ChatGPT stands out for the quality of its representations and a greater propensity to suggest mitigation strategies or technological improvements. Perplexity obtains equally valid results, often with a comparable level of technical detail; however, the graphic component (such as tables and layouts) is sometimes less accurate and appealing compared to ChatGPT.

As an illustrative example, Figure 3 shows how chatbots can accurately perform a sector-specific analysis focused on the technical aspects of the devices under examination, integrating information extracted from the provided documentation with insights gathered through online research.

Feature	Garmin Edge 1050	Garmin Edge 840	Garmin Edge 540	Wahoo ELEMNT Roam
Display Type	3.5" Color Touchscreen	2.6" Color Touchscreen	2.6" Color Non-Touch	2.7" Color Non-Touch
Resolution	282 × 470 px	246 × 322 px	246 × 322 px	240 × 400 px
GPS Type	Multi-band GNSS (GPS, Galileo)	Multi-band GNSS	Multi-band GNSS	GPS, GLONASS, Galileo
Battery Life	Up to 20h (45h with battery saver)	Up to 26h (42h with battery saver)	Up to 26h (42h with battery saver)	Up to 17h
Storage	32GB Internal	16GB Internal	16GB Internal	No Internal (Cloud Sync)
Wi-Fi	✅	✅	✅	✅
Bluetooth & ANT+	✅	✅	✅	✅
LTE	❌	❌	❌	❌
Navigation	Turn-by-turn, ClimbPro, Trailforks, Heatmaps	Turn-by-turn, ClimbPro, Trailforks	Turn-by-turn, ClimbPro, Trailforks	Turn-by-turn, Breadcrumb Navigation
Smart Features	Garmin Pay, Music, Phone Notifications	Phone Notifications, Strava Sync	Phone Notifications, Strava Sync	Phone Notifications, Strava Live Segments
Safety Features	Incident Detection, LiveTrack, GroupRide, Emergency Contacts	Incident Detection, LiveTrack, GroupRide	Incident Detection, LiveTrack, GroupRide	LiveTrack, Crash Detection
Power Meter Support	✅	✅	✅	✅
ClimbPro Feature	✅	✅	✅	❌
Indoor Trainer Support	✅ (Smart Trainer Control)	✅	✅	✅
Weight	160g	85g	80g	93g
Price Range	$750 - $850	$500 - $650	$430 - $540	$400 - $500

Figure 3: Example of competition analysis generated by chatbots (technical features of the examined devices)

Overall, both chatbots confirm their effectiveness as tools for business analysis, capable of collecting and reworking information from multiple sources, offering comprehensive summaries and strategic insights useful for the development of new products, significantly reducing the workload of human analysts.

5. Conclusion

This research focused on use and effectiveness of generative chatbots within business contexts, with particular attention to their potential for business analysis and decision support. As the referenced literature, while emphasizing the growing interest in GenAI, highlights the lack of concrete application guidelines to assist companies in adopting these technologies, this paper aims to offer an initial contribution in this direction, presenting a study that addresses both practical and methodological aspects.

The results highlight that GenAI tools, such as ChatGPT and Perplexity, can significantly streamline processes, simplify information management, and make strategic analyses more timely (see research question RQ1 in the Introduction). The evidence collected demonstrates the ability of these chatbots to support companies in collecting and interpreting data, thereby providing useful insights to guide decisions and complement traditional and human-based analysis (as per RQ2).

However, critical issues remain unresolved, particularly the difficulty in managing certain file formats, the variability of responses based on the specific context (when this context is not fully explained to the system), and the need for constant human intervention to validate the outputs produced by AI (RQ3). This necessitates users' ability to input appropriate prompts, interact with the system to refine questions and data inputs, and interpret the results. In other words, while chatbots are intended to assist managers in their decisions, human action cannot be entirely replaced, and new capabilities are required to use these systems effectively.

A significant aspect of the study was the comparison between ChatGPT and Perplexity, which revealed notable differences in their approach to data and accuracy of results. This comparison is particularly important given the variety of technologies and tools available on the market, each with distinct performances and characteristics. In our case, ChatGPT demonstrated excellent processing and overall reliability, although it faced challenges with complex tabular data. Perplexity, on the other hand, provided more detailed answers but showed some limitations in accuracy.

From an operational perspective, the results suggest that the use of generative chatbots could lead to greater efficiency in decision-making processes, thanks to the speed of analysis and the ability to support decisions. To maximize the benefits, a collaborative model in which AI is supported by professionals who validate and interpret the system's outputs still appears advisable (RQ4).

The implications of this research for theory include advancing our understanding of AI's role in KM and decision support, while for practice, it provides insights into the practical applications and limitations of GenAI tools in business contexts, and underscores the importance of developing user guidelines and best practices for the effective and responsible adoption of GenAI technologies.

Looking to the future, there are numerous areas for development. Improvements in reading structured formats, the creation of shared methodologies for evaluating chatbot performance, and integration with other business intelligence tools are just a few. Additionally, future research could explore the ethical implications of AI in decision-making, the development of more robust frameworks for AI-human collaboration, and the impact of AI on organizational culture and employee roles.

The paper has some limitations that should be considered. First, the study was based on a structured experimental plan that included over 210 different experiments and approximately 900 prompts but, due to length constraints, only a summary is provided in this paper. Second, further experiments should be conducted with different cases and business applications. Third, the replication of experiments may suffer from the subjective intervention of researchers, who might use different variations of prompts to interact with the chatbots. Additionally, part of the assessment criteria used to evaluate results are subjective in nature, which may introduce biases. Nonetheless, we believe this exploration can provide valuable insights and pave the way for new lines of research on how chatbots can contribute to producing a useful knowledge base for business decisions, a critical topic in KM literature.

Acknowledgements

This study was funded by the project “Fourteen” financed by the University of Padova.

AI Declaration: GenAI Chatbots (ChatGPT and Perplexity) were used to conduct the experiments that are integral part of the empirical study. A GenAI chatbot (Lucrez-IA), developed by the University of Padova, was used to check English of some specific sentences.

Ethics Declaration: No ethical clearance was required for this study.

References

Agrawal, A., et al. (2022) "Exploring the Capabilities and Risks of Large Language Models", arXiv preprint arXiv:2207.04504.

Alavi, M., Leidner, D. and Mousavi, R. (2024) "Knowledge management perspective of generative artificial intelligence (GenAI)", *Alavi, Maryam*, pp 1-12.

Arrieta, A.B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... and Herrera, F. (2020) "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI", *Information fusion*, Vol 58, pp 82-115.

Cano-Marin, E. (2024) "The transformative potential of Generative Artificial Intelligence (GenAI) in business: a text mining analysis on innovation data sources", *ESIC Market*, Vol 55 No. 2 pp e333-e333.

Cillo, P. and Rubera, G. (2024) "Generative AI in innovation and marketing processes: A roadmap of research opportunities", Journal of the Academy of Marketing Science, pp 1-18.

Csaszar, F.A., Ketkar, H. and Kim, H. (2024) "Artificial intelligence and strategic decision-making: Evidence from entrepreneurs and investors", Strategy Science, Vol 9 No. 4 pp 322-345.

Duan, Y., Edwards, J.S. and Dwivedi, Y.K. (2019) "Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda", International Journal of Information Management, Vol 48, pp 63-71.

Feuerriegel, S., Hartmann, J., Janiesch, C. and Zschech, P. (2024), "Generative AI", *Business & Information Systems Engineering*, Vol 66 No. 1 pp 111-126.

Gichoya, J.W., Thomas, K., Celi, L.A., Safdar, N., Banerjee, I., Banja, J.D., ... and Purkayastha, S. (2023) "AI pitfalls and what not to do: mitigating bias in AI", *The British Journal of Radiology*, Vol 96 No. 1150, 20230023.

Jarrahi, M.H., Askay, D., Eshraghi, A. and Smith, P. (2023), "Artificial intelligence and knowledge management: A partnership between human and AI", *Business Horizons*, Vol 66 No. 1 pp 87-99.

Kanbach, D.K., Heiduk, L., Blueher, G., Schreiter, M. and Lahmann, A. (2024), "The GenAI is out of the bottle: generative artificial intelligence from a business model innovation perspective", *Review of Managerial Science*, Vol 18 No. 4 pp 1189-1220.

Kirchner, K., Bolisani, E., Scarso, E., Kassaneh, T.C. and Taraghi, N. (2024) "Integrating Generative AI into Knowledge Management: The Case of Software Development", In *Proceedings IFKAD 2024* (pp.386-398).

Kolbjørnsrud, V., Amico, R., & Thomas, R. J. (2016). How artificial intelligence will redefine management. Harvard Business Review, No. 2, pp. 1-6.

McKinsey and Company (2023) "The State of AI in 2023", QuantumBlack AI, McKinsey, August 2023, Retrieved from https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2023-generative-ais-breakout-year

McKinsey Global Institute (2012) "The Social Economy: Unlocking Value and Productivity Through Social Technologies, Retrieved from https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/the-social-economy

Morgan Stanley Research. (2025) "GenAI Revenue Could Surpass $1 Trillion by 2028", Retrieved from https://www.morganstanley.com/insights/articles/genai-revenue-growth-and-profitability

Nakash, M. and Bolisani, E. (2024) "Knowledge Management Meets Artificial Intelligence: A Systematic Review and Future Research Agenda", In *European Conference on Knowledge Management* (pp.544-552), Academic Conferences International Limited.

Nakash, M. and Bouhnik, D. (2024) "How Much Time does the Workforce Spend Searching for Information in the 'new normal'?", *iConference 2024 Proceedings*.

Shrestha, Y. R., Ben-Menahem, S. M., & Von Krogh, G. (2019). Organizational decision-making structures in the age of artificial intelligence. *California management review*, Vol. 61, No. 4, pp. 66-83.

Soni, V. (2023) "Adopting generative ai in digital marketing campaigns: An empirical study of drivers and barriers", Sage Science Review of Applied Machine Learning, Vol 6 No. 8 pp 1-15.

Storey, V.C. (2025) "Knowledge management in a world of generative AI: Impact and Implications", *ACM transactions on management information systems*.

Stray, D. (2021) "AI for the Workplace: Hype, Skills, and Limitations", IEEE Software, Vol 38 No. 4 pp 10-16.