EARLY WARNING SYSTEMS AGAINST BANKRUPTCY RISK AND NLP: CAN CHATGPT PREDICT CORPORATE DISTRESS?

Article · June 2023

DOI: 10.29119/1641-3466.2023.170.28

1 author:

Jędrzej Siciński
University of Gdansk

SILESIAN UNIVERSITY OF TECHNOLOGY PUBLISHING HOUSE

SCIENTIFIC PAPERS OF SILESIAN UNIVERSITY OF TECHNOLOGY 2023
ORGANIZATION AND MANAGEMENT SERIES NO. 170

EARLY WARNING SYSTEMS AGAINST BANKRUPTCY RISK AND NLP: CAN CHATGPT PREDICT CORPORATE DISTRESS?

Jędrzej SICIŃSKI

University of Gdańsk, Faculty of Management, Department of Strategic Development; Sopot; jedrzej.sicinski@ug.edu.pl, ORCID: 0000-0001-6810-3792

Purpose: The main purpose of this paper is to evaluate the effectiveness and usability of one of the more groundbreaking and more widely commented NLP-technique-employing inventions, i.e., the ChatGPT application acting as a digital advisor in the field of counterparty financial standing and bankruptcy risk assessment.
Design/methodology/approach: The algorithmic potential presented by the ChatGPT tool can be a valuable solution in supporting the manager's work. In this study, the current potential of this solution in supporting financial analysis, and in particular, bankruptcy risk assessment, was checked. The study was carried out using the following methods: analysis and synthesis (1.), critical analysis of the literature (2.), and an experiment involving the use of a natural language processing application (3.).
Findings: In the course of the research, it was found that the ChatGPT tool, according to the current state of knowledge, has extensive usability and is able to conduct interactions that in many cases are similar to communication with a human being. The tested language model shows a much higher level of training on general data than in solving narrow problems in specific fields. Nevertheless, its development potential should be assessed highly and probably its adaptation to solve highly specialized tasks in management will not be a long-term process, which makes it a candidate for the role of a digital managerial advisor in the future.
Research limitations/implications: The first stage of the research covered only solving problems with the use of the simplest algorithms, such as discriminant analysis (MDA) and the study of entities whose financial statements are widely available on the web, which was a relatively low level of complexity for the language model.
Practical implications: The research results are a signal that digitization and the digital revolution are not just theoretical slogans, but real functioning technologies that can change the nature of the manager's work (and the entire management system) in the near future. The development potential of NLP technology in management, which was confirmed in this work, suggests that an appropriate strategy for implementing these technologies is needed today.
Originality/value: In this study, one of the first attempts was made to assess the potential and adaptability of natural language processing systems to support a manager in assessing the financial condition and risk of bankruptcy of entities.

Keywords: early warning systems against bankruptcy, insolvency, artificial intelligence, natural language processing, chatGPT.

Category of the paper: Research paper, Literature review.

1. Introduction

Artificial intelligence technology is being implemented in an increasing range of areas of daily human life. Spanning from transportation flow support (Paprocki, 2017), through strategies of marketing message individualization, to cell phones, the more or less complex algorithms consistently attempt to predict user behavior and environment changes, simultaneously improving themselves. Artificial intelligence solutions are also increasingly widely applied in management – by relieving cadres of simple, repetitive tasks, they allow for, inter alia, process automation (Sliż, 2019) or detection of fraud and abuse, based on vast data sets (Aslam et al., 2022). It should be emphasized the efficiency thereof is often many times higher than that of the same activity carried out by a human, while the entire work is often free of defects and errors that are inherent in human work. An increasing amount of discussion nowadays is devoted to the involvement of artificial intelligence in more creative management tasks, such as consulting or support during managerial decision-making at not only operational but also tactical and strategic levels. One important aspect to stress entails the fact that the mere implementation of artificial intelligence in an enterprise does not necessarily bring spectacular business results if it is not preceded by development of a well-thought-out plan and a strategy for the implementation thereof (Sira, 2022). The development of an implementation plan is essential to identifying the areas most susceptible to AI support in a particular enterprise. In management, one of the areas utilizing numerical algorithms and artificial intelligence, characterized by a long tradition of AI application, is the so-called early warning against corporate bankruptcy (Siciński, 2021). Using historical data derived from financial statements, these measures enable enterprise classification and thus identification of entities at risk of bankruptcy, several years in advance. This allows appropriate anticipatory reaction on the part of a manager, such as e.g., a change in the trade credit policy towards such an entity or discontinuation of business cooperation. Early warning algorithms can be built using both simple mathematical methods as well as complex, sophisticated artificial intelligence and machine learning methods (Shetty, 2022). Since early warning models offer significant support in risk management (e.g., in the formation of counterparty credit policies) (Antonowicz, 2007), managerial cadres have for years shown interest in the implementation of such models in daily management. The main barrier to the implementation and everyday use thereof, however, still lies in the fact that even the simplest EWS (Early warning systems) require certain analytical and financial skills on the part of the users, as well as prior acquisition of quality financial data. Even more advanced early warning systems using e.g., machine learning or artificial intelligence are extremely difficult to utilize in practical management, as they often take the form of a so-called black box, i.e., they do not offer any separate, user-visible functions with parameters. This analytical form renders these systems extremely unintuitive for managers lacking data-science expertise. The issue of how to make classification and prediction methods more accessible and understandable to mass users in management practice is not covered broadly in the literature. The research deficit particularly concerns the realm of bankruptcy early warning systems. This particular issue represents a research gap identified by the Author, which needs to be filled. In the Author's view, one prospective approach to systems of early warning against bankruptcy, which can contribute to the wider and more effective use thereof in management, entails natural language processing (NLP) technology. Such an approach can be referred to as a so-called hybrid technique, which involves the use of artificial intelligence support in handling other data-science techniques that are beyond the reach of managers lacking adequate knowledge. Similar attempts to support the users are already underway, e.g., using external software (a robot) by means of voice commands, artificial intelligence can generate complex computational functions in an MS Excel spreadsheet, with the ability to export those functions into the spreadsheet.

The purpose of this paper is to evaluate the effectiveness and usability of one of the more groundbreaking and more widely commented NLP-technique-employing inventions, i.e., the ChatGPT application acting as a digital advisor in the field of counterparty financial standing and bankruptcy risk assessment. This assessment was carried out using so-called systems of early warning against the risk of bankruptcy. The study was aimed at verifying whether the ChatGPT algorithm has knowledge of the essence and mathematical specification of as well as the rules of application and inference from the most popular bankruptcy risk assessing models.

The research process planned is of a long-term nature and will be materialized with a series of publications, whereas the research results presented in this article are preliminary and concern the results of the ChatGPT algorithm's interaction with the least complex early warning technique (often of first choice) - namely, multiple discriminant analysis (MDA). The research questions posed were as follows:

RQ1. Do natural language processing algorithms exhibit utility in the application of other data-science-based solutions in the sphere of finance and early warning against insolvency?

RQ2. Does the current effectiveness of advanced NLP methods indicate that in the future, specialized chat-bots will be able to comprehensively assess the financial condition of a counterparty and recommend appropriate management actions, reducing thereby the need for external expertise and work-intensive analyses?

The following were employed as the research methods: analysis and synthesis (1.), critical analysis of the literature (2.), and an experiment involving the use of a natural language processing application (3.). The research hypothesis posed states: Intelligent natural language processing algorithms can provide support to managers as digital advisors in assessing the external standing of a company (e.g., a contractor), using systems of early warning against bankruptcy.

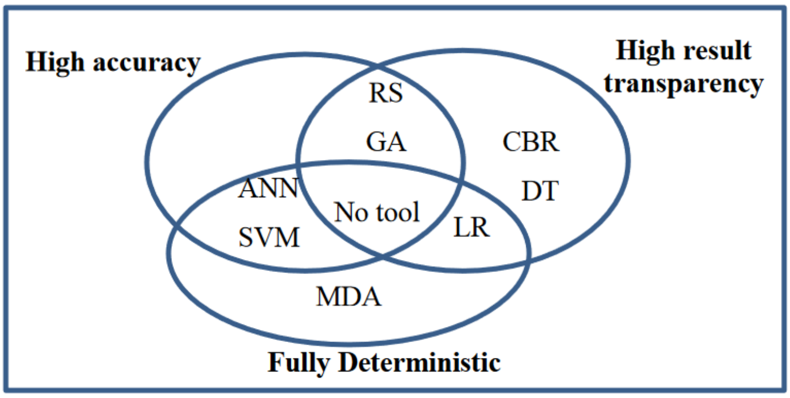
2. Early warning systems against bankruptcy – literature review

Early warning systems (EWS) are designed to determine the risk of a company's bankruptcy with sufficient time in advance (minimum one year). Such a model can take the form of a linear equation (sometimes a complex algorithm), the purpose of which is to dichotomously classify a company using a point predictive score based on selected data contained in the financial reports issued by the company. It is believed that one of the first historical events that led economists to consider the possibility of predicting financial risk was the Great Depression of the 1930s (Kowalak, 2017). Over the following decades, the development of such models accelerated, mainly as a result of two factors: the improvement of data analysis methods (1.) and the increasingly broader and effective corporate financial reporting obligation (2.) (Siciński, 2021). A distinction can be made between three main quantitative methodological currents, in which models of early warning against bankruptcy are developed (Pociecha et al., 2014):

- discriminant analysis,
- econometric methods,
- artificial intelligence methods (including neural networks).

In practice, other approaches, embedded in experimental and higher mathematics, such as entropy theory and fuzzy logic, can also be distinguished. Moreover, machine learning methods, such as random forest and Extreme Gradient Boosting (XGBoost) algorithms, are gaining popularity as well. The latter (XGBoost), which has been under test in bankruptcy risk prediction environment for a relatively short time, yet achieves very high predictive performance, seems to be a particularly promising one (Zięba et al., 2016). Regardless of the wide range of methods, the Z-Score early warning system developed by Edward Altman (Altman, 1968) is considered in the financial world as a pioneering, most popular model.

The multiplicity of bankruptcy risk assessment methods and techniques should be deemed positive, in the context of this trend's development and further advancement. Nevertheless, typically, the predictive accuracy of these methods increases with the sophistication of the algorithm, and this in turn leads to increased difficulty of use for recipients, who accordingly require correspondingly higher knowledge of financial analysis and data science. It can thus be argued that the above-mentioned methods of early warning model construction are characterized by three fundamental user attributes, which, as a rule, exert adversarial effects on one another. This means that, in principle, the high deterministic nature of functions entails lower predictive accuracy and lower transparency of results, and vice versa - neural networks are characterized by lower determinism and low transparency yet high efficiency of bankruptcy risk prediction (Alaka et al., 2018). This relationship is shown in Figure 1.



Note. MDA - Multiple Discriminant Analysis, LR - Logistic regression, ANN - Neural Network, SVM - Support vector machines, RS - Rough Sets, CBR - Case Based Reasoning, DT - Decision Tree, GA - Genetic Algorithm.

Figure 1. Performance of tools in relation to result-related criteria.

Source: Alaka, H.A., Oyedele, L.O., Owolabi, H.A., Kumar, V., Ajayi, S.O., Akinade, O.O., Bilal, M. (2018). Systematic review of bankruptcy prediction models: Towards a framework for Tool Selection. Expert Systems with Applications, 94, 164-184. https://doi.org/10.1016/j.eswa.2017.10.040.

The very application of fully deterministic methods, which are considered the least complicated (e.g., MDA - multivariate discriminant analysis or LR - logit regression), requires basic knowledge of the accounting, financial and data-analysis principles. This is because assessment of bankruptcy risk by means of such models involves acquisition of data, calculation of relevant metrics and subsequent determination of the function's scoring, along with its interpretation. Moreover, discriminatory functions are, in principle, characterized by a very high sensitivity to changes in input parameters, and even seemingly insignificant errors in the calculation of metrics can contribute to radical changes in the result of the final classification. Even experienced analysts not infrequently can belittle a fact of an erroneous approach to the proper averaging of certain financial categories (a requirement for some financial analysis indicators) (Antonowicz, 2011), which can distort the results of insolvency risk assessment.

Summing up, systems of early warning against bankruptcy are characterized by a long history of use, while the model itself can be built to varying methodological standards. These methods differ in effectiveness as well as the scale of difficulty in application and interpretation by the user (e.g., a manager).

3. Natural language processing (NLP) in ChatGPT algorithm

Natural language processing (NLP) is categorized as one of the so-called cognitive technologies and is often customarily classed within the boarder category of artificial intelligence (Osika, 2021). NLP can be defined as a set of various computational techniques that are oriented toward analysis of text and human-like understanding thereof (Liddy, 2001). This means that natural language processing models human characteristics by giving computers the ability to respond to various commands delivered through spoken and written language (IBM, 2023). The technology owes its capabilities to sophisticated computational techniques, which employ mathematical models, statistics, machine learning and so-called deep learning. NLP has a wide range of practical applications, the most widespread forms of which include (Khurana et al., 2022):

- text translation,
- text categorization,
- support of anti-spam systems,
- establishment of interactions in the form of dialogues (e.g., through so-called chat-bots),
- sentiment analysis.

This technology, as evident by the above examples of application, touches almost every person, oftentimes in ways they are unaware of. We come into contact with NLP even through sending and receiving e-mail, for instance, where a relevant system ensures reduction of unwanted advertising message (SPAM) influx. Likewise, interaction with most online stores or any form of off-premises product acquisition is no different - online sales often involve human interaction with bots guiding customers through the sales or complaint process. Nevertheless, the greatest state-of-art NLP technology invention is, first and foremost, the GPT-3 language model (Maciąg, 2022) and the hugely gaining media popularity language processing system called ChatGPT. The effectiveness of comprehension and interaction, as well as the overall representation of human nature in these models, are sufficiently high for the models to be attributed with very significant impact on social life and the world around us (Tamkin et al., 2021). The effectiveness of such a language model can be assessed using the so-called Turing test (Turing, 1950). The technique involves judges (real users) engaging in a conversation with an unseen interlocutor. If, during a 5-minute conversation, even a share of those engaged in the dialogue (minimum 30% of the judges) are unable to distinguish whether they have interacted with a bot or a real person, such an algorithm should be considered as meeting the requirements of the Turing test (Gov.pl, 2023). Research into ChatGPT is still ongoing, but unofficial reports indicate that it is likely capable of passing the Turing test procedure (Mpost, 2023). The outstandingly high efficiency in user interaction demonstrated by ChatGPT in various applications is also contributive to the discussion on the potential role of this tool in financial management. Cognition of the model's potential in solving not only predictable and linear problems (these issues are already pursued freely in business practice) but complex matters requiring creativity and abstract thinking specifically, is of particular importance.

4. Experiment methods and assumptions

The study employed three research methods: analysis and synthesis, critical analysis of the literature, and an experiment with the use of an algorithm processing natural language (ChatGPT). The assumptions underlying the ChatGPT experiment carried out took the form of a four-phase scenario. Three discrimination functions of early warning against insolvency (one of foreign and two of Polish origin) were selected for this purpose:

1) E.I Altman's discriminant function (Z-Score) model:

$$\mathrm{Z} = 1{,}2\mathrm{X}_1 + 1{,}4\mathrm{X}_2 + 3{,}3\mathrm{X}_3 + 0{,}6\mathrm{X}_4 + 0{,}999\mathrm{X}_5 \quad (1)$$

where:

X1 – net working capital/total assets,
X2 – retained earnings/total assets,
X3 – ebit/total assets,
X4 – enterprise value/book value of debt,
X5 – sales/total assets.

2) Prusak's P1 discriminant function model:

$$\mathrm{Z}_{\mathrm{P1}} = 6{,}5245\mathrm{X}_1 + 0{,}1480\mathrm{X}_2 + 0{,}4061\mathrm{X}_3 + 2{,}1754\mathrm{X}_4 - 1{,}5685 \quad (2)$$

where:

X1 – ebit/total assets,
X2 – operating costs/short-term debt,
X3 – current assets/short-term debt,
X4 – ebit/sales.

3) INEPAN's[1] Z7 discriminant function model (developed by a scientific team led by Elżbieta Mączyńska):

$$Z_{7INE} = 9{,}498X_1 + 3{,}566X_2 + 2{,}903X_3 + 0{,}452X_4 - 1{,}498 \quad (3)$$

where:

X1 – ebit/assets,
X2 – total equity/total assets,
X3 – earn after tax + depreciation)/total debt,
X4 – current assets/short-term debt.

[1] INEPAN: The Institute of Economics of the Polish Academy of Sciences [Instytut Nauk Ekonomicznych Polskiej Akademii Nauk] is a public Polish research center for economic and business studies, with rights to bestow doctoral and habilitation degrees as well as initiate professorship procedures.

The inclusion of models estimated by the discriminant analysis method (MDA) in the experiment was motivated by the fact that the technique has been most widely used (the so-called ‘first choice’ method). The article additionally presents the first stage of the Author's current research in this domain, hence the selection of this method as the first to interact with the ChatGPT algorithm can be considered legitimate. The functions selected for the study are considered to be among the most reliable and accurate in the practice of financial risk prediction (Antonowicz, 2007) and are frequently featured in various analyses within this research stream (Kitowski, 2021). The assumptions underlying individual phases of the experiment, including the related goal of knowledge exploration and objective to assess the language model’s potential with respect to the issue under study are summarized in Table 1.

Table 1.
ChatGPT language model experiment assumptions - four phases of the interaction scenario

Phase	Model’s interaction issues	Phase Objective
1.	General knowledge of early warning systems against bankruptcy	Checking whether the language model has knowledge of the basic assumptions underlying systems of early warning against bankruptcy.
2.	Area knowledge of specific bankruptcy models	Checking whether the model is equipped with knowledge on mathematical specification of selected discriminant functions.
3.	Calculation of discriminant function score	Assessment of the language model’s analytical and mathematical skills and its ability to search for accurate financial enterprise data.
4.	Formulation of conclusions regarding the risk of insolvency	Checking whether the model can formulate correct conclusions based on the results of each discriminant function.

Source: Own elaboration.

Proper execution of all four interaction scenario phases will indicate the full capability of the ChatGPT language model in handling the most popular methods of discriminant analysis in early warning against bankruptcy. Correct implementation of only the first two phases will indicate the algorithm's theoretical knowledge of bankruptcy early warning systems, with absence of utility potential as managerial support in the sphere of finance. Moreover, a situation where theoretical and general knowledge of early warning systems is low (Phases 1 and 2) is also possible, and yet the ChatGPT algorithm is still capable of making correct calculations and accurate inferences from the mathematical function. Such a result shall be referred to as an unconscious skill, i.e., correct use of the early warning system, without a proper theoretical layer. In the third and fourth phases, fifteen randomly selected companies, whose shares are publicly traded and included in the main indices of selected world stock exchanges, were incorporated into the study. The collective encompassed: ten companies listed on the Warsaw Stock Exchange (WSE), ten listed on the German stock exchange (DAX), and ten on the US stock exchange (Dow Jones). The primary language of communication with the ChatGPT algorithm was Polish. The proposed size of the collective should be considered sufficient, since the architecture of language models (as well as that of other artificial intelligence solutions) exhibits a dichotomous approach. If the algorithm shows ability to solve a certain problem on object i_1, it is generally also able to solve the same problem on object i_2.

5. Results

The results of the user interaction with the ChatGPT model are presented in accordance with the experimental plan outlined earlier (Table 1). Phase one and phase two ended with a positive or negative result. A positive result in the theoretical stages (phases one and two) occurred when the following conditions were jointly met:

- the messages generated by the language model combined logic with proper argumentation,
- the response generated by the chatbot allowed sufficient implementation of the preformulated phase objective (Table 1),
- the ChatGpt responses provided were of characteristics similar to natural human communication.

With respect to the third and fourth phases (calculation stages), the totals of entities classified correctly after entering the command "Calculate the index ....(Altman, P1 Prusak, Z7 INE PAN) for company X in 2021" into the chatbot are tabulated in the corresponding Table 2 cells. More so, in case of an incorrect answer to this phrase, a second attempt was undertaken using a modified command "Determine the 2021 bankruptcy risk for entity X by means of model X." Ultimately, after executing the commands for the three discriminant functions and all the companies listed in the indices selected, the language model was assigned a final overall score on its ability to use each function for bankruptcy risk estimation. Definitive positive assessment resulted when the language model was able to execute a minimum of one theoretical phase correctly and obtain a minimum of 50% of accurate classifications and conclusions based on the computational phases carried out for the exchange entities selected. The results of each phase are summarized in Table 2.

Table 2.
Results of four-phase interaction with natural language processing model (ChatGPT)

Model / **Experiment phase (stage)**	**Z-Score discriminant function (E.I Altman)**	**P1 Prusak discriminant function**	**Z7 INE PAN discriminant function**
Results from the phase 1	Positive		
Results from the phase 2	Positive	Negative	Negative
Results from the phase 3	Number of correct calculations of the scoring function:		
WSE	7/10 (70%)	0/10 (0%)	0/10 (0%)
DAX	10/10 (100%)	0/10 (0%)	0/10 (0%)
DOW JONES	10/10 (100%)	0/10 (0%)	0/10 (0%)
Results from the phase 4	The number of correctly formulated conclusions from the result of the scoring function:		
WSE	7/10 (70%)	0/10 (0%)	0/10 (0%)
DAX	10/10 (100%)	0/10 (0%)	0/10 (0%)
DOW JONES	10/10 (100%)	0/10 (0%)	0/10 (0%)
Overall score of ChatGPT in each discriminant function	Positive	Negative	Negative

Source: Own elaboration.

The first phase tested the ChatGPT model's general knowledge of bankruptcy early warning systems. The linguistic algorithm responded correctly to the inputted command *Do you know what a system of early warning against bankruptcy is?* In the feedback message, the model formulated a logical and concise answer, stating it is a system aimed at early detection of financial risks before a company's situation becomes too serious for implementation of corrective measures. The quality of this message should be considered sufficient and satisfactory; thus, the first phase can be assessed as successful. The only drawback entails the fact that the language model failed to pinpoint the most important attribute of EWS solutions, namely that bankruptcy risk assessment results from the processing of indicators calculated from financial data. Phase two went far less successfully. A command was introduced into the tool: *Specify the formula of the discriminant function '...'* (the names of the functions checked for were entered in the designated ellipsis, respectively: E.I Altman / Z7 INE PAN / P1 Prusak). The model was able to correctly describe and show the specification of E.I Altman's (1968) discriminant function but had very limited knowledge of the leading bankruptcy models designed for Polish economic conditions. With respect to the Altman Z-score, the ChatGPT algorithm formulated the model specification, number of variables, names of the variables, and the values and signs of the structural discriminant function's parameters correctly. It also accurately indicated the so-called cut-off points and the theoretical rule for inferring the result. When dealing with INEPAN's Z7 and Prusak's P1 models, the language model confirmed that it was familiar with these solutions, yet it formulated the functions incorrectly, misstated the number and names of the discriminating variables, and improperly composed the numerical assessments of the structural parameters. Occasionally, it also confused these solutions with completely different financial analysis tools.

In the calculation phases (stage 3 and 4), the language model operated very well with the E.I Altman's discriminant function. It correctly calculated the score for all selected DAX and DOW Jones index listed entities, as well as carried out 70% of positive calculations for WSE-listed companies. It is noteworthy that correct calculations were also followed by well-formed conclusions regarding the risk of bankruptcy. This means that the ability to correctly calculate a given function was always accompanied by the ability to correctly interpret the result. Referring to the leading discriminant functions developed in Poland (Prusak's P1 function and INEPAN's Z7), practical knowledge application (calculation phases) was fully correlated with the lack of theoretical knowledge on the two (phase 2). The linguistic algorithm failed to respond adequately to the commands regarding determination of the discriminant function value for a given company, although it correctly drafted and presented selected financial data for potential calculations, explaining that its knowledge in this area was running out. The algorithm, however, suggested risk assessment for the company in question using another bankruptcy model it was familiar with. This phenomenon highlights a degree of information noise which probably occurred during the ChatGpt algorithm training – in the large datasets processed, the model had most likely come across theoretical information on the models

(Z7 INE PAN, P1 Prusak) but blended it with information on other, unspecified discriminant functions from unverified sources. This presumably indicates an under-training of the algorithm with data from outside the United States and/or information recorded in a language other than English, as well as its tendency to cross-reference disparate information. The current revision of the ChatGPT model, according to the developers' assertions, is based on a training dataset dated at the end of 2021. This means that interactions regarding tasks, events and facts occurring after that time are significantly limited. The two models (Z7 INE PAN, P1 Prusak) are among the most widely used ones in Poland and have been the subject of several hundred publications in both Polish and English (with numerous citations), but they were developed several years ago, which is a serious indication of the language model's non-time-related deficiencies. This implies that there is a large number of potential input (training) data, both spatially and temporally, which the algorithm probably did not process properly in the learning process. Referring still to E.I Altman's discriminant function, the algorithm furthermore rightly signaled its limitations, stressing: *the Altman index comes with certain limitations and does not necessarily accurately reflect the financial situation of a company* [translated from the original in Polish by author].

6. Discussion

The results of the study carried out suggest a noticeable under-training of the algorithm when it comes to the realm of non-U.S. data, particularly if the interaction with the language model involves complex, specific issues and requires prior acquisition of secondary financial data. Nevertheless, the algorithm shows impressive training at the level of general issues, fulfilling, so to speak, the role of a ‘communicating encyclopedia.’ Referring to its management support suitability, specifically as a digital managerial advisor in the assessment of corporate financial health or bankruptcy risk, the algorithm is characterized (according to the current state of knowledge) by a clearly inferior potential to infer and formulate abstract judgments with additional input data requiring further processing (e.g., corporate financial reports). In other words, the model is almost flawless in its analytical operations, when it receives an entire set of input data in a command from the user, as opposed to when it is constrained to acquire this data on its own. This should not be considered a categorical disadvantage, as the process was to some extent intended by the developers of the tool - the model is meant to gain a broad multi-domain communication capability of building on huge datasets (WEF, 2023) rather than specialize in a narrow problem area. The current potential of the ChatGPT language model (and other language models currently being developed by competitors), however, indicates that its adaptation to more specialized and industry-specific problems should not be particularly challenging. What is more, steps are already being taken in the business world to implement natural language processing systems into management support, which is in line with the projections of the Author of this article. One example is the project developed by a well-known venture capitalist Peter Thiel, aimed at development of a managerial chat-bot supporting financial directors and managers in day-to-day accounting and finance decision making (Bloomberg, 2023). According to the Author of this paper, the key to developing such a solution (or to specialized/industry-wide adaptation of such models as ChatGPT) is to provide enough high-quality data, which will serve as a training set, and integrate the tool with a structured digital database of financial reports (e.g., Bisnode). The combination of these two factors, should enable construction of a chatbot characterized by intuitiveness and freedom of communication similar to the ChatGPT model, but enriched with highly specialized knowledge of finance, financial statements, and bankruptcy risk analysis. Such a combination would be the likely seed for establishment of a fully functional digital financial management advisor for executives.

7. Summary

The digital revolution, inextricably accompanied by the expansion of artificial intelligence, has been stirring up many extreme emotions in society, ranging from the hope that machines will completely free the human species from tedious, repetitive tasks, to more radical expectations that all of humanity will soon be able to stop working and robots will be taxed, which will lead to technological unemployment (Moser, 2021). This will give every human being the right to a permanent income in the form of a so-called technological dividend, payable regardless of status, and financed by taxing robot labor. These predictions coincide with the fears expressed in the past, more specifically, in the lead-up to the Great Industrial Revolution. Indeed, at that time, fears were expressed that machines and the associated mass production would deprive the society of jobs. Nowadays, opinions can be found that intelligent solutions, including artificial intelligence, will wipe out many of today's well-known and practiced professions. In such borderline considerations, the truth usually lies in the middle. One highly probable scenario is that within 10-20 years, every human engaged in professional work will become, to a greater or lesser extent, a so-called ‘robot shepherd’ (Forsal, 2023). This should be interpreted to coincide with the actual effects of the 18^{th} century industrial era - most workers did not lose their jobs altogether, but simply became machine operators. A similar scenario is materializing today in many areas of the economy - in an ever-increasing number of professions, workers who were once fully dependent on analog work tools must coordinate and supervise the operation of intelligent systems, often merely fulfilling the role of an automated process operator. Referring to the ChatGPT model’s potential in supporting human labor, especially in the process of financial management, however, it often exhibits characteristics coinciding with the ‘Renaissance mind’. The tool, in its current version, efficiently handles interactions across nearly any subject sphere - from social sciences to higher mathematics (provided it receives adequate information input and assumptions from the user). It should be borne in mind, though, that this is still a solution trained on general knowledge, as well as on Internet resources, the quality of which is not fully controlled, as evidenced by the results of the experiment and the errors identified at the level of even the second phase of interaction. Any mathematical model training its algorithm with dirty data, for instance, will, to some extent, carry these distortions over onto the final prediction results, leading to the so-called ‘GIGO’ (Garbage in, garbage out) effect (Najman, Migdał-Najman, 2018). The study conducted seems to confirm that the current version of and the degree to which the language model implemented in the ChatGpt tool has been trained does show impressive features at a general level, but specialized issues requiring it to launch an exploration of Internet resources (e.g., independent acquisition and selection of relevant financial annual report data) often fall beyond its capabilities. This is true even for publicly traded entities, whose financial reports are widely available in various formats in the Web resources, and should again be attributed to the nature of the invention, as it was trained with a dataset predisposing it to resolving general knowledge problems rather than to the narrow realm of financial engineering and bankruptcy risk. Nevertheless, it ought to be emphasized that, based on the results of individual experiment phases, the model probably holds unlimited development potential and its adaptation to the issues of bankruptcy risk assessment and the broadly understood business standing assessment, should the need arise, would be a matter of months rather than years. Even now, financial institutions are successfully employing simple, autonomous installment loan granting systems, while the increasingly widespread marketing of BNPL (Buy now, pay later) solutions is fully based on artificial intelligence. More earnest and creative attempts to adapt chatbots to managerial support are likewise taking place, as evidenced, inter alia, by projects under development by well-known VC investors. The research questions formulated should thus be answered as follows:

RQ1: Natural language processing algorithms (including ChatGPT), even at this stage, demonstrate utility in the implementation and application of such other data-science-based methods as e.g., bankruptcy risk assessment models.

RQ2: The current efficiency in text processing by NLP-type systems, combined with the intuitiveness thereof for the user suggests that in the near future, many tasks involving the broadly defined assessment of corporate financial health (including counterparty bankruptcy risk) may be automated by intelligent Chatbots. This trend's existence is evidenced not only by the skills demonstrated by the ChatGPT model in the experiment, but also by the extensive commercial attempts to create ‘management advisor chatbots’, identified in the course of literature research.

Referring to the research hypothesis posed, it should be affirmed that intelligent natural language processing algorithms are capable of contributing managerial support in the assessment of a company's standing (e.g., using systems of early warning against bankruptcy). It should be caveated, however, that the capability is strongly conditioned by access to an appropriate dataset on which the algorithm will be ‘trained’ and highly dependent on proper integration of the tool with a database of high-quality financial data. The trend to refine these tools in this area of expertise is positively developmental, which means that soon many managerial activities will be able to receive autonomous support. In the foreseeable future, will an opportunity arise for an appropriately trained bot or virtual assistant processing natural language to respond adequately to the question *Check whether company X is a reliable contractor and whether I should agree to credit a sale to this entity, and if so, under what conditions?* The current state of knowledge and the results of the research conducted suggest that the answer is "yes". Certainly, the direction for further research lies in the realm of how ChatBot-type solutions will cope with more advanced financial threat prediction tools (e.g., when operating with black box tools, i.e., without a visualized function).

The main limitations in the study carried out stem from the fact that the experiment was based on existing and ready-made discriminant functions, which not infrequently show differential predictive performance, depending on the entity sector and country of origin. This means that the language model was confronted with a relatively undemanding task - namely, the need to use a ready-made, universal function to assess bankruptcy risk. The direction of further research, and at the same time a serious challenge (which currently likely lies beyond the potential of the Chatgpt model), would be to use an optimally selected classification function (e.g., matched to the sector or life cycle stage of the entity under investigation), or have the NLP model design its own function or neural network in real time at the user's command, each time it is queried about a company's bankruptcy risk/financial health. This would certainly require a redefinition of the entire solution's architecture, although in this case, the algorithm would not need to be trained on a broad, general data set (which customarily is very time-consuming), but only on information of a narrower and specialized nature (financial risk and financial analysis). This kind of NLP model capability (to create custom AI solutions in real time) would then bear traits of next economic revolution predicted by futurologists (after the coming digital and artificial intelligence revolution), namely the era defined by the moment when Artificial Intelligence begins to produce its own AI (robots produce their own robots). Indeed, a similarity can be discerned - a natural language processing algorithm, as a robot, designs an autonomous neural network (another robot), tailored and trained to the nature of the inputted command, to maximize the prediction results (e.g., a particular entity’s financial problem/risk of bankruptcy).

References

1. Alaka, H.A., Oyedele, L.O., Owolabi, H.A., Kumar, V., Ajayi, S.O., Akinade, O.O., Bilal, M. (2018). Systematic review of bankruptcy prediction models: Towards a framework for Tool Selection. *Expert Systems with Applications, 94*, pp. 164-184. https://doi.org/10.1016/j.eswa.2017.10.040
2. Altman, E.I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *Journal of Finance, 23(4),* pp. 589-609. https://doi.org/10.1111/j.1540-6261.1968.tb00843.x.
3. Antonowicz, P. (2007). *Metody oceny i prognoza kondycji ekonomiczno‑finansowej przedsiębiorstw.* Gdańsk: Wydawnictwo ODiDK.
4. Antonowicz, P. (2011). Długość cyklu inkasa należności i rotacji zapasów w przedsiębiorstwach na trzy lata przed sądowym ogłoszeniem upadłości. *Problemy Zarządzania, no. 1(31),* pp. 53-71
5. Aslam, F., Hunjra, A.I., Ftiti, Z., Louhichi, W., Shams, T. (2022). Insurance fraud detection: Evidence from Artificial Intelligence and machine learning. *Research in International Business and Finance, 62,* 101744. https://doi.org/10.1016/j.ribaf.2022.101744
6. Bloomberg (2023). *CFOs Will Get Their Own AI Chatbot as Part of Brex Project.* Retrieved from: https://www.bloomberg.com/news/articles/2023-03-07/brex-to-offer-chatgpt-style-openai-chatbot-for-cfos-finance-professionals#xj4y7vzkg?leadSource=uverify%20wall, 09.03.2023.
7. Forsal (2019). *Pasterz robotów i manager śmierci cyfrowej. W ciągu 10 lat powstaną zupełnie nowe zawody*. Retrieved from: https://forsal.pl/artykuly/1410248,praca-przyszlosci-w-ciagu-10-lat-powstana-zupelnie-nowe-zawody.html, 07.03.2023
8. GOV.PL (2020). *Test Turinga i sztuczna inteligencja*. Retrieved from: https://www.gov.pl/web/ncbr/sztuczna-inteligencja--czy-pamietamy-o-tescie-turinga, 27.02.2023.
9. IBM (2023). *Natural language processing.* Retrieved from: https://www.ibm.com/topics/natural-language-processing, 27.02.2023.
10. Khurana, D., Koli, A., Khatter, K., Singh, S. (2022). Natural language processing: state of the art, current trends and challenges. *Multimedia Tools and Applications, 82(3),* 3713-3744. https://doi.org/10.1007/s11042-022-13428-4.
11. Kitowski, J. (2021). Modele dyskryminacyjne jako instrument oceny zagrożenia upadłością przedsiębiorstw. *Nierówności Społeczne a Wzrost Gospodarczy, 68(4),* 145-160. https://doi.org/10.15584/nsawg.2021.4.8.
12. Kowalak, R. (2017). Rola systemów wczesnego ostrzegania w procesie podejmowania decyzji o restrukturyzacji przedsiębiorstwa. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu, no. 470,* p. 159.

13. Liddy, E.D. (2001). *Natural Language Processing*. Encyclopedia of Library and Information Science, 2nd Ed. NY. Marcel Decker, Inc. p. 2.
14. Maciąg, R. (2022). Zaawansowane procedury NLP jako przesłanka rekonstrukcji idei wiedzy. *Zarządzanie w Kulturze, 23(1),* pp. 37-53.
15. Migdał-Najman, K., Najman, K. (2018). Dirty data – profiling, cleansing and prevention. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu, 508,* 146-156. https://doi.org/10.15611/pn.2018.508.15.
16. Moser, E. (2021). Against robot taxes: scrutinizing the moral reasons for the preservation of work. *AI And Ethics, 1(4),* 491-499. https://doi.org/10.1007/s43681-021-00062-3.
17. MPost (2023). *Chaptgpt passes the turing test*. Retrieved from: https://mpost.io/chatgpt-passes-the-turing-test/, 27.02.2023.
18. Osika, G. (2021). Dilemmas of social life algorithmization - technological proof of equity. *Scientific Papers of Silesian University of Technology Organization and Management Series, 151.* https://doi.org/10.29119/1641-3466.2021.151.36.
19. Paprocki, W. (2017). Perspektywy zastosowania sztucznej inteligencji w systemach obsługi mobilności - przykład organizacji pasażerskich przewozów lotniczych. *Research Journal of the University of Gdańsk. Transport Economics and Logistics, vol. 74,* pp. 341-355.
20. Pociecha, J., Pawełek, B., Baryła, M., Augustyn, S. (2014). *Statystyczne metody prognozowania bankructwa w zmieniającej się koniunkturze gospodarczej*. Kraków: Fundacja Uniwersytetu Ekonomicznego w Krakowie, p. 8.
21. Shetty, S., Musa, M., Brédart, X. (2022). Bankruptcy Prediction Using Machine Learning Techniques. *Journal of Risk and Financial Management, 15(1),* 35. https://doi.org/10.3390/jrfm15010035.
22. Siciński, J. (2021). *System wczesnego ostrzegania przedsiębiorstw przed ryzykiem upadłości na przykładzie branży transportowej*. Sopot: Centrum Myśli Strategicznych, pp. 8-10.
23. Sira, M. (2022). Artificial Intelligence and its application in Business Management. *Scientific Papers of Silesian University of Technology. Organization and Management Series, 165,* 307-346. https://doi.org/10.29119/1641-3466.2022.165.23.
24. Sliż, P. (2019). Robotization of business processes and the future of the labor market in Poland–preliminary research. *Organizacja i Kierowanie, 185(2),* p. 67-79.
25. Tamkin, A., Brundage, M., Clark, J., Ganguli, D. (2021). *Understanding the Capabilities, Limitations, and Societal Impact of Large Language Models,* ArXiv:2102.02503. Retrieved from: https://arxiv.org/abs/2102.02503v1, 27.02.2023.
26. Turing, A.M. (1950). Computing Machinery and Intelligence. *Mind, 236(14),* 433-460.

27. WEF (2023). *How does chagpt differ from intelligence.* Retrieved from: https://www.weforum.org/agenda/2023/02/how-does-chatgpt-differ-fromhuman-intelligence?utm_source=facebook&utm_medium=social_scheduler&utm_term=Artificial+Intelligence&utm_content=28%2F02%2F2023+20%3A40, 7.03.2023.
28. Zięba, M., Tomczak, S.K., Tomczak, J.M. (2016). Ensemble boosted trees with synthetic features generation in application to bankruptcy prediction. *Expert Systems With Applications, 58,* 93-101. https://doi.org/10.1016/j.eswa.2016.04.001.