The application of artificial intelligence in corporate finance

By

Quinten Kropmans (s2886774)

Master Business Administration
University of Twente

2024

Mentored by M. Huiskes

1st Supervisor: Dr. W. Van Heeswijk

2nd Supervisor: E. Svetlova

Abstract

Despite several researches conducted about Large Language Models (LLM) and their applications in practice, there were no empirical researches conducted about the applications of LLMs in corporate finance Mergers & Acquisitions (M&A). Research has been conducted about the application of AI in M&A, but on different aspects of the process like the mandatory Due Diligence process in M&A. This research was conducted at the corporate finance company Taurus. In the process of M&A, an Information Memorandum (IM) is written to inform potential buyers about every important company topic. This IM includes an internal and external analysis of the company. The internal analysis contains the key characteristics of the company itself. The external analysis is an analysis of the company's market and contains important information that is of interest to a potential buyer. These analyses require a lot of valuable time from M&A consultants. Therefore, this research aims to find out how LLMs can make the writing of internal and external analyses more efficient. The research was conducted by interviewing three corporate finance experts and an AI expert and assessing LLM-generated documents according to the evaluative analysis method. The interviews were conducted to gain a broader understanding of the IM's structure and identify important assessment variables using the Gioia method. The assessments were executed by using an Assessment Format, that was formulated using evaluative analysis and the coding scheme. To assess whether LLMs can be used in M&A, we recreated IMs using Gemini and ChatGPT 4o. The IMs that were recreated are original IMs that Taurus experts wrote for actual M&A processes. This was done for 5 different IMs. Gemini scores higher on every aspect, compared to ChatGPT. When looking at the overall assessments of the LLM-generated texts, Gemini receives a score of 66 compared to a score of 46 for ChatGPT. This is a major difference of 20 points out of a total score of 100. LLMs are very useful for internal and external analysis of companies and can make this process more efficient. However, the output always needs to be reviewed and adjusted when necessary. If Taurus decides to adopt my suggestions, employees at Taurus will need to learn how to work with LLMs and gain experience, and eventually, the writing process at Taurus will change substantially, as writing texts can largely be taken over by LLMs. The scope for this research opens up 3 future research directions to focus on. One important direction for future research is a more broad research design, where more assessments will be taken into account to create a more substantiated study.

Table of Contents

1. Introduction

Corporate finance is a specific area of finance. It involves the strategies that corporations use for financing and investment decisions to maximize their value, addressing a wide range of risks and leveraging financial reports for decision-making. It also encompasses corporate governance, focusing on how corporate boards act and set company values and balancing stakeholder interests for success (Gherghina, 2021). Corporate finance is crucial in every business, as it accompanies financial decisions and strategies companies implement. Professionals in this field, for example, financial analysts, CFOs, and investment bankers play important roles in helping businesses make well-considered and substantiated financial decisions that align with the company's strategic objectives. In corporate finance, several sub-fields can be separated, such as capital budgeting, capital financing, and working capital management (Mang'ana, Hokororo, & Ndyetabula, 2024). A sub-field of corporate finance is Mergers & Acquisitions (M&A). M&A is the process of selling and purchasing companies. In this research, the focus will be in the field of M&A only.

M&A is a very dynamic and complex process. Mainly because of this complexity, it is often a time-consuming process. This is why companies involved in M&A processes, usually engage a specialist to assist them during this process. Taurus Corporate Finance (Taurus), located in Deventer in The Netherlands, is such a specialist focusing on M&A advisory. The services Taurus provides are the assistance of entrepreneurs in a process of buying or selling a company, company valuation, and the restructuring and recovery of companies in dire straits. This research will be conducted at the company Taurus. The scope for this research will be the M&A process since this is the most important service Taurus provides. The M&A process from a corporate finance perspective can involve either the buy side or the sell side. Taurus focuses primarily on sell-side deals. The personal interaction between the advisor and sellers of the company is crucial. The advisor and those involved at the selling company will work closely together to create an Information Memorandum (IM) that fits well with a realistic representation of the company sold. Customization is necessary due to the complex and unique nature of financial issues that companies may encounter. However, new opportunities to simplify and automate these processes may be available in the form of artificial intelligence (He Wang, 2023). Taurus desired research to be conducted, to find out what the opportunities are for the application of AI within their sell-side advisory operations.

Artificial intelligence, commonly known as AI, has emerged as a revolutionary force in our rapidly evolving technological landscape, reshaping how we interact with machines, process information, and even understand human intelligence (Ergen, 2019). This technology can transform entire markets, boost productivity, and open up new opportunities in various industries, from healthcare and banking to entertainment and transportation (Cao, 2022). AI refers to the development and deployment of computer systems or algorithms that possess the ability to perform tasks that typically require human intelligence (Shi & Zheng, 2006). As AI develops, its applications expand in scope and sophistication, affecting both the commonplace and remarkable facets of our lives (Dwivedi, et al., 2021). Specific applications of AI include expert systems, natural language processing, speech recognition, and machine vision (Craig, Laskowski, & Tucci, 2024). This introduction offers a glimpse into the complex world of AI by examining its fundamental ideas, practical uses, and revolutionary effects on how we understand and use machine intelligence (Shao, Lou, Wang, Mao, & Ye, 2021). A form of AI that is upcoming in the rapidly changing world that we currently live in, are large language models (LLMs). LLMs, such as the Generative Pre-trained Transformer (GPT), have made significant advancements in natural language processing in recent years. LLM represents AI tools that are trained on vast amounts of data to generate human-like text, answer questions, and complete other language-related tasks with high accuracy (Alberts, et al., 2023; Kasneci, et al., 2023). There are a lot of public LLMs, with the most well-known LLMs being ChatGPT and Gemini. The difference between traditional language models and ChatGPT and Gemini, is that traditional language models are programmed to use statistical techniques to predict the next word in a sentence, whereas, ChatGPT and Gemini use transformer-based models that allow for the processing of vast amounts of data in parallel (Alberts, et al., 2023). LLMs could be used to formulate a text for the analysis. Currently, it is unclear how reliable LLM-generated text is. Unclarity about the reliability of texts can make it difficult to trust and use the information effectively. This research needs to be conducted, to find out how LLMs can facilitate the M&A process, and how reliable the results of the analysis are. AI systems are widely accepted as a technology offering an alternative way to tackle complex and ill-defined problems (Kalogirou, 2003). They can learn from examples, are fault tolerant in the sense that they can handle noisy and incomplete data, can deal with non-linear problems, and once trained can perform prediction and generalization at high speed (Kalogirou, 2003). AI has been used in diverse applications in control, robotics, pattern recognition, forecasting, medicine, power systems, manufacturing, optimization, signal processing, and social/psychological sciences (Kalogirou, 2003). AI is particularly useful in system modeling such as in implementing complex mappings and system identification (Kalogirou, 2003).

Currently, it is not possible for AI to fully replace human consultants (Harding, D'Alessandro, Laskowski, & Long, 2023). The owner of a selling company often experiences a lot of emotions when they are selling their company (Tikkakoski, 2018). An important task for the M&A consultant in sell-side advisory, is to work with those emotions and realize the best possible deal for the owner of the company. Every owner values different aspects when selling their business. These personal preferences are currently not possible to take into account for LLMs. The best possible deal does not mean the deal that is generating the highest amount of money. What the best deal is, depends on the preferences and values of the seller. Some sellers prefer the highest prices, while others want the company to stay intact, after the transaction. An M&A consultant's task is to help the owner with these decisions. LLMs cannot decide what would be the best decision for the seller, based on feelings, conditions and desires of this seller. This research will be conducted by interviewing Taurus experts and an AI experts. Additionally, Taurus experts will assess LLM-generated texts. Based on both the interviews and the assessments, a well-considered conclusion could be drawn how LLMs can assist and improve efficiency in the generation of internal and external analysis documents. This will be possible because the AI expert will provide us with all important and underlying information from LLMs and what the application could do for Taurus. The interviews with the Taurus experts will help to give a broad view of how IMs are built and what LLMs can do to assist in the writing process of IMs. The LLMs can analyze information very quickly. Human beings would not be able to compete with this speed. LLMs can generate applicable analyses if the right prompt is used for the right purpose. The best way to generate a prompt for Taurus' analysts will be investigated and analyzed in Chapter 3. Because LLMs are a lot faster than human beings, this will increase the efficiency of the writing process. However, an LLM cannot understand context which frequently results in disinformation (Chen & Shu, Combating misinformation in the age of LLMs: Opportunities and challenges, 2024). This is one of the biggest risks in using LLMs. To mitigate the risk of using disinformation, every generated text needs to be reviewed on various aspects like accuracy, completeness, correctness, sources, and documents received from the company. However, it is also possible for an LLM to generate a text and create an analysis, that the corporate finance expert has not accounted for. This can be useful for consultants to get new insights and inspiration during a project (Asadi, 2023, December 26).

The focus for this research will be on the internal and external analysis of companies that are customers at Taurus for accompaniment for M&A sell-side advisory. The internal analysis of a company will inform potential buyers about all the key characteristics of this particular company. A sophisticated internal analysis will prevent a buyer from being faced with unwanted surprises in the process. The external analysis is an analysis of the market that the company operates in. A potential buyer will be informed about potential threats, opportunities, technologies, and other important topics. Those analyses are used for pitches and IMs. Even though the analysis per company is the same in general, it needs to be customized for each company, which is very time-consuming. An external analysis for a company also needs to be customized, because one company fits in an environment differently than another company. For example, a tax advisory firm operates in a comparable environment as an accountancy firm, but its core business is different. Other aspects of a company matter as well, like size and geographical presence. An internal and external analysis usually consists of a short text, containing between 500 and 6,500 words to explain the subject, and a few visuals to enhance the textual explanation. The number of words depends on the complexity and size of the analyzed company. Larger and/or more complex companies require more explanation than smaller and/or less complex companies. The internal and external analysis in the IM are both divided into smaller subchapters. Information that will be explained in these subchapters are the strategy of a company, trends in the market, the customers and suppliers, competitors, housing of the company, and personnel for example.

Taurus is curious how LLMs could be used to improve the efficiency and speed of these analyses. The current problem with the internal and external analysis is that the process takes a lot of time to execute (per company). LLMs could be a solution to reduce the time that would be required to write an internal and external analysis (Cockburn, Henderson, & Stern, 2019; Pallathadka, et al., 2023). The purpose of the research is to demonstrate whether LLMs could be applied to assist in this process. Since LLMs are particularly good to use for writing texts, it could be interesting to investigate if certain tasks that require a text to be written could be taken over by LLMs. When the writing part for the internal and external analysis of a company could be taken over by LLMs, with the same quality and reliability, then this would result in a much quicker analysis and process. Currently, there is not a lot of research conducted on the application of LLMs in analysis, and this research is to figure out what the benefits of LLMs could be.

1.1 The M&A Process at Taurus

This M&A process starts with a couple of meetings between the consultants of Taurus and the shareholders of the company that is intended to be sold. The sellers engage Taurus by signing a mandate letter. A Deal team to assist with this project will be selected, which varies from 2 to 5 consultants, depending on the complexity of the project. After the first meeting, Taurus' deal team will send out an Information Request List (IRL) to help the sellers to select information to be submitted to a secure online database. This document contains specific information to make a thorough analysis of the company's financial, legal, and tax aspects. This analysis will be made available to any prospective buyer by means of an IM. Draft versions of the IM will be made available to the sellers and management team in order to check and double-check the completeness and correctness of the information. The IM needs to represent the company in an honest and best possible way. In the final version of the IM, a process letter will be added, which contains the timelines for interested parties in the acquisition process. This starts with the preconditions for a non-binding indicative offer (NBIO). After Taurus has received all the NBIOs from the potential buyers, the negotiation phase will start. The preparation phase, visualized in Figure 1, is the phase during which the IM is written. This is usually the longest phase in terms of time.

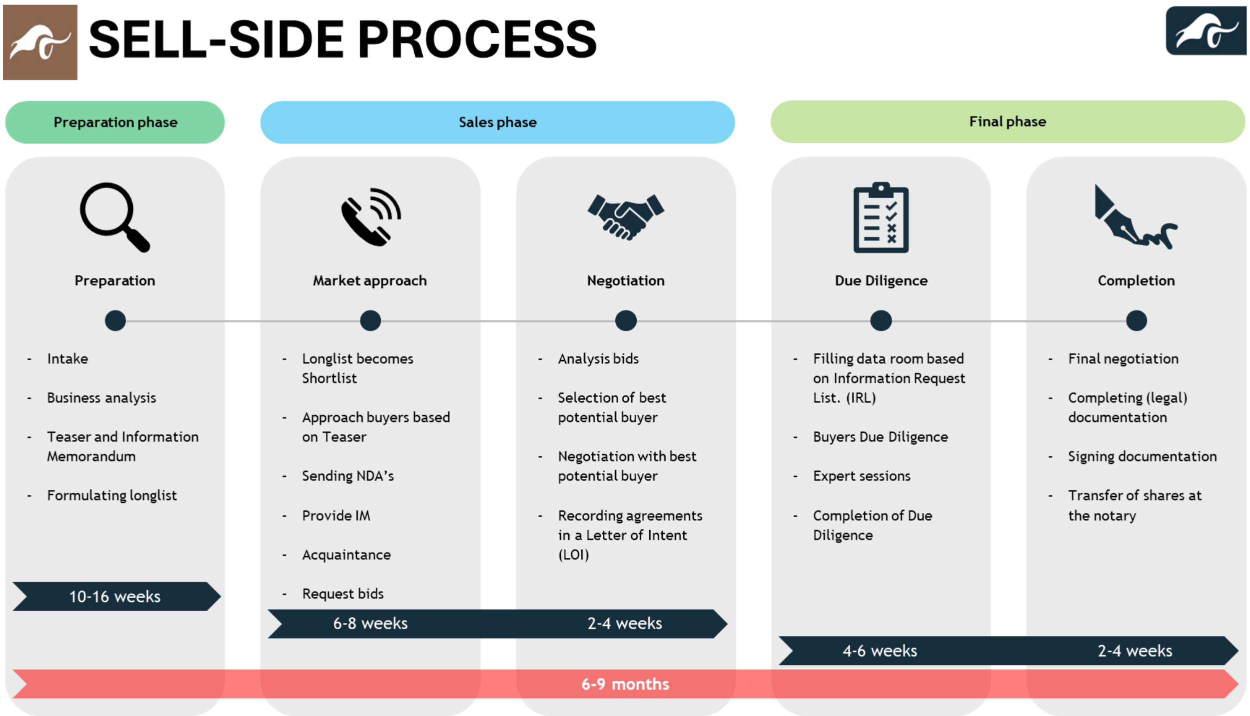

Figure 1. *An overview of the sell-side process. The overview visualizes how the M&A process at Taurus roughly works.*

The purpose of using LLMs in this process is to make the process more efficient. During the IM writing process, the deal team simultaneously develops a list of potential buyers, based upon certain criteria, e.g. industrial fit, growth rate, investment strategy, ticket size, national/international, majority/minority shares, and potential need for new management. The potential buyers will be selected by Taurus, where the potential buyers can be both strategic and financial. A strategic buyer could be a competitor, and a financial buyer could be a private equity party. In a management meeting, this list of potential buyers will be discussed. New names will be added and proposed names will be rejected (e.g. due to insufficient funds or negative connotations of the seller). A short list of approximately 10 companies/investors will be contacted by Taurus.

To make sure that the selling process will remain confidential, the potential buyer needs to sign a non-disclosure agreement (NDA). After signing the NDA Taurus will send out the IM and the approached parties are invited to make an NBIO. The most attractive bidders will be invited to a management meeting. In this meeting, the bidder will have the full opportunity to ask additional questions and information and meet the management team in person. The Taurus' consultant will contact this potential buyer afterwards to invite them to eventually revise their initial offer. All revised offers will be made comparable by the consultant and the seller will select the most attractive bid. With this potential buyer, Taurus will negotiate a letter of intent (LOI). After signing the LOI the bidder will be invited to have their LOI preconditions reviewed in a due diligence process (DD). The IM and the management meetings should reflect a rightful image of the company. External advisors of the buyer will perform this DD and advise the client on the factual state of the company and the expected cash flows. This might lead to re-negotiations on the conditions of the deal, but at Taurus, the comprehensive pre-deal research proves valuable. Where other corporate finance companies make significant concessions during negotiations, Taurus is less willing to compromise. This is because the IM is constructed in the most honest, realistic, and reliable possible way. 80% of the deals will be closed as laid down in the LOI. After confirmation, the legal transaction documents can be drafted. This is not only a share purchase agreement, but on many occasions, also a vendor loan agreement, a management agreement, and a shareholder agreement (if the seller is to reinvest partially).

1.2 Problem Description

Currently, the preparation phase of the M&A process at Taurus is quite extensive. The purpose of this research is to find out if the process can be made more efficient using LLMs. This master thesis investigates how AI, and specifically LLMs, can contribute to a more efficient flow within an M&A process. The purpose of this research is also to evaluate whether the LLMs are reliable. By using LLMs, labor-intensive and repetitive tasks could be automated, which not only could improve efficiency but also could increase the accuracy, quality, and speed within this section of the process. This thesis will analyze and evaluate the potential application of LLMs in the field of M&A, to determine the added value of this technology for corporate finance consultants and companies in terms of efficiency. This will be done by interviews and assessments. Three interviews will be held with Taurus experts, along with one interview with an AI expert outside of Taurus. The LLMs will generate texts that will be assessed by six different experts at Taurus. Two of those six experts were interviewed initially as well.

1.3 Research Question

Since LLMs are a relatively new technology, there is not a lot of research executed on these applications. To the best of our knowledge, this study is the first that examines how capable LLM can execute specific analyses in an M&A process. The purpose is to find out what the capabilities of LLMs are in the internal and external analysis of several companies, and if the LLM-generated texts are qualitatively similar to texts written by Taurus experts. This will be investigated in trying to answer the research question:

How can the LLMs Gemini and ChatGPT 4o make the internal & external analysis of companies, that are performed in the process of Mergers & Acquisitions at Taurus Corporate Finance, more efficient?

The theoretical contribution of this research to the literature would be that experts in the field of corporate finance M&A will have a foundation to understand how LLMs can assist them while making an internal and external analysis of a company. If the LLM-generated texts are useful, then the application of those texts could save a lot of time for Taurus. To judge whether the texts are qualitatively similar, the experts from Taurus will assess the LLM-generated texts and the IMs that experts at Taurus already wrote.

The practical contribution of this research would be that LLM-generated texts can be implemented more efficiently in the IMs that Taurus experts write. This research investigates what kind of input you have to provide to the LLMs, to generate appropriate LLM texts. It would help Taurus to work more efficiently, and save time so that the consultants can focus on other tasks within the M&A process.

The academic contribution of this research would be the scientific substantiation, to decide how to prompt the LLMs ChatGPT and Gemini, to receive useful results when it comes to an internal and external analysis of a company.

The following sub-questions will jointly answer the research question in various chapters:

1. **What is already known about AI and LLMs and its applications in an M&A process? (Chapter 2)**
2. **What research method is best to compare the original analysis document and LLM-generated analysis in terms of relevancy, clarity, structure, and overall quality? (Chapter 3)**
3. **How to formulate queries and prompts to make the LLMs generate an accurate text, that could be used in an Information memorandum? (Chapter 4)**
4. **Which LLM is the most suitable to improve the process of generating analytical reports of companies, and to improve the efficiency of the analysis for Taurus? (Chapter 5)**

2. Theoretical Background

The impact and application of LLMs in corporate finance, especially the M&A process, and the effect that it could have on the speed and quality of analysis are poorly understood. The impact on finance in general is better understood. LLMs can enhance financial AI by improving natural language processing tasks but require careful selection based on data, computing, and performance constraints (Li, Wang, Ding, & Chen, 2023). Chapter 2.1 will provide an overview of the literature on corporate finance and M&A. Chapter 2.2 will briefly explain what LLMs are, and Chapter 2.3 will dive deeper into LLMs in M&A. The world of M&A is dynamic and integration of LLM within this landscape could enhance decision-making processes. The rise of AI could announce a new era of accelerated analysis and augmenting the quality of decision-making within M&A.

2.1 Mergers & Acquisitions

This research explores the theoretical substantiations that relate to M&A. Companies can have very complex ownership structures, that could delay the M&A process (Nogueira & de Castro, 2019). Additionally, it can take a long time to complete because of the complexity of the process, and because small changes in the process can have an impact on the final bid for a company (as described in Chapter 1.1.1). Therefore, it is important to go through every step of the process carefully. This process is very complex and risky with a high failure rate and is deeply influenced by the sector to which the organizations belong (del Río Carballo, González, & Sumaza, 2023). During the M&A process, in almost every case, an IM and Teaser will be written to inform potential buyers about the company's aspects (Gomes, 2018). This IM provides information about, for example, the history of a company, its personnel, its real estate, the market that the company operates in, possible opportunities and threats, and other important information (Gomes, 2018). Between financial institutions and companies, M&A have become very common due to various worldwide factors (Darayseh & Alsharari, 2023).

2.2 Large Language Models

2.2.1 Large Language Models Definition

LLMs are a form of AI that is capable of generating texts (Head, Jasper, McConnachie, Raftree, & Higdon, 2023). Because AI is a far too broad topic, for this research we will only focus on

LLMs and their capabilities. LLMs are trained on large datasets and they operate on machine learning and natural language processing (Yao, et al., 2024). They can create large amounts of text in a short period. However, one of the limitations of LLMs is that the texts generated are not always based on actual sources. The nature of LLMs is that it provides plausible-sounding answers based on a Transformer model which is a type of deep neural network (Vaswani, et al., 2017). A Transformer model is what powers both Gemini and ChatGPT. LLMs show promising progress in AI because of this Transformer model, enabling models trained to predict the next word in a text to perform other tasks with intelligence (Douglas, 2023). An LLM predicts what the most likely next word will be in a text, based on the input that the LLM has been provided with, and what the LLM already generated (Yamauchi, Sonoda, Sannai, & Kumagai, 2023). In an aim for human-like artificial general intelligence, LLMs are a major advancement (Arcas, 2022).

2.2.2 Large Language Models in Practice

LLMs are used in various fields for a wide range of tasks, including law, to automate tasks like legal judgment prediction, document analysis, and writing (Sun, 2023; Yang, et al., 2024). They are increasingly recognized for their potential to transform various aspects of work, for example revolutionize legal tasks, but their integration faces challenges like privacy concerns, bias, and explainability (Sun, 2023). However, they can also offer capabilities that can increase efficiency, enhance productivity, streamline communication, and support decision-making in various fields of work (Bellomo, Zhang, Ivers, Cohen, & Ozkaya, 2023, November). LLMs can especially improve productivity by automating routine tasks (Rasnayaka, Wang, Shariffdeen, & Iyer, 2024). LLM-based tools show promising results in scientific research, but their effectiveness varies across use cases and output integrity remains a concern (Nejjar, Zacharias, Stiehle, & Weber, 2023). In many areas of today's work life, LLMs have been demonstrating increased productivity (Nejjar, Zacharias, Stiehle, & Weber, 2023). A lot of studies have been executed about the applications of LLMs in medical practices. In the medical sector, LLMs look very promising but require validation and supervision due to variations in accuracy and incorrect outputs (Sorin, et al., 2023). However, it is unclear how LLMs can increase productivity in the area of M&A. Nevertheless, it is important to keep in mind that other research showed that the LLM-generated texts require validation and oversight (Ryu, et al., 2023, December; Hennigen, et al., 2024). This indicates that LLMs could be useful as a tool within M&A, but not as an unconditional replacement of an employee. (Zhao, et al., 2024) Investigated the application of LLMs in finance. In this research, LLMs are only mentioned in the field of M&A in the financial forecasting of an M&A transaction (Zhao, et al., 2024). They use LLMs to analyze different variables and possibly predict a transaction (Zhao, et al., 2024). This research does not go into the content of the M&A process itself (Zhao, et al., 2024). However, if the use of LLMs can save sufficient time, a company could work with fewer employees, or do more projects with the same number of employees.

2.3 LLMs in M&A

Other research shows how the development of AI is being used in data analysis in the M&A process, and what the advantages and disadvantages are of the implementation of AI in the M&A process (Wang & Zhou, 2023). AI is used in the due diligence process in M&A, but its consequences on safety and ethics are showing up more (Li Y. , 2023). The disadvantage of traditional analysis is that it takes human judgment, it is time-consuming and labor-intensive (Wang & Zhou, 2023). Human people tend to make more mistakes, compared to AI or machines (Wang & Zhou, 2023). A dependence on human resources increases the likelihood of errors (Wang & Zhou, 2023). This study shows the implementation of AI in 3 subprocesses of M&A, Due Diligence, Negotiation and Post-Merger (Wang & Zhou, 2023). The cons of implementing AI in M&A are: increasing cyber security risks, people losing their jobs because of the implementation of AI, and to keep AI functional it is time-consuming and expensive (Wang & Zhou, 2023). Both (Rien, 2018; Bedekar, et al., 2024) Explore the application of AI in the Due Diligence process of M&A. Due Diligence is a complex process requiring a lot of time and is mandatory to close an M&A process. Out of all M&A transactions, about 70% are unsuccessful or fail to create value (Martin, 2016). The purpose of a Due Diligence process is helping reduce the risk of unsuccessful M&A transactions (Rien, 2018). The application of AI in the Due Diligence process is investigated more often. Currently, no detailed empirical study about the use of LLMs in internal and external analysis of companies in an M&A process has been executed. Therefore, this research must be conducted, to better understand how LLMs can assist in the M&A process. However, there have been a variety of researches conducted about the use of AI in the different stages of the M&A process. This research is unique because it empirically studies the use of LLMs in internal and external analysis of companies in an M&A process, which has not been done in the past. Other research has shown that strategy management analyses can be generated by LLMs for the most part (Brath, Bradley, & Jonker, 2024). The purpose of this strategy management analysis is not to be used in an M&A process. Therefore, this research does not fully connect with our research.

3. Methodology

This chapter outlines the research design and the way the experiment will be executed to investigate how LLMs can make the internal and external analysis of companies more efficient, by using various prompts to generate texts that are comparable to the original material, as stated in Chapter 1.3. The materials that will be used are the analyses that the Information Memoranda (IMs) contain. During the process of M&A, an internal and external analysis needs to be conducted to give a clear and complete image of the company that is being sold to a potential buying company. As previously described in Chapter 1.2, the analyses consist of a short text to explain the subject, and a few visuals to enhance the textual explanation. As described in Chapter 1.2, these visuals are out of scope for this research, and the focus will be on the textual description of the companies. LLMs are made to write texts based on the prompt that a person provides to the system. This research will evaluate the capabilities of LLMs for M&A activities and the best application for this technology in this field of work. The research will focus on the internal and external analysis of companies. The research will be conducted by comparing existing analysis with LLMs-generated analysis. By using an assessment format (Appendix A) that was formulated according to the coding scheme of the interviews (Appendix D), the best LLM for the IM will be chosen. The results will show which LLM would generate the most applicable texts for the internal and external analysis, when the LLMs are prompted with publicly available sources.

3.1 Quantitative or Qualitative research

When it comes to text analysis, quantitative and qualitative analysis are two fundamental types of research. They each have their distinct methodologies, applications, and purposes. Quantitative text analysis is the application of methods for drawing statistical inferences from text populations, using instrumental, representational, and network-related approaches (Roberts, 2000). Quantitative analysis involves mainly the collection and analysis of numerical data. This research method often involves statistical tools to test hypotheses, identify patterns, and make predictions. In contrast, qualitative analysis focuses on understanding events through non-numerical data in the form of texts. Qualitative data analysis involves examining the data to understand its meaning and implications, and identifying patterns and trends (Seers, 2011). Qualitative data analysis is a systematic and strict approach that aims to answer questions about what something is like, what people think or feel about it, and why it has happened (Seers, 2011). Common methods of executing qualitative analysis are conducting interviews, content analysis, focus groups, and ethnography (Bristowe, Selman, & Murtagh, 2015). Both quantitative and qualitative analyses are essential in research. Quantitative analysis is used for studies requiring measurable evidence and statistical outcomes. Qualitative analysis is valuable for contextual factors and human experiences. The purpose of the research is to demonstrate whether ChatGPT and Gemini can be used in the M&A process at Taurus. Also, it is important to figure out how reliable and usable the LLM-generated output is. Because LLMs generate texts, this does fit better in the qualitative research methodology, as it allows deeper exploration of linguistics and contextual meaning (Kuckartz, Qualitative Text analysis: a Systematic approach., 2019; Davis, 1995). Quantitative analysis will most likely overlook this exploration. Qualitative analysis facilitates a better understanding of emphasis of certain topics within the text. Results in a text are not easily quantifiable, as the data is non-numerical (Kuckartz, Qualitative Text analysis: a Systematic approach., 2019). Qualitative research in applied linguistics emphasizes individuality and context, while quantitative research focuses on large populations and ignores individual differences (Benson, 2012). In this research, we do not have large populations. Therefore, considering the information mentioned above, the best method to execute this research is through a qualitative analysis, since the LLM-generated texts need to be judged by experts.

3.1.1 Qualitative Text Analysis

A qualitative text analysis needs to be executed to accurately compare the LLM-generated text with the text written by Taurus experts. This way the experiment will give the most accurate conclusion to the research question. Qualitative text analysis generally has three basic methods: thematic analysis, evaluative analysis, and type-building analysis (Kuckartz, Three basic methods of Qualitative text analysis, 2014). Thematic analysis is used for analyzing qualitative data and it is one of the most commonly used (Kuckartz, Qualitative Text analysis: a Systematic approach., 2019). Often Thematic analysis is called Qualitative Content Analysis (QCA) in Europe (Kuckartz, Qualitative Text analysis: a Systematic approach., 2019). This method is used to analyze verbal and textual data, ensuring rigor and consistency in the analysis, and therefore, QCA is a systematic, rule-based process (Schilling, 2006). QCA is used within data using conventional, directed, or summative approaches to identify, analyze, and report patterns and themes (Braun & Clarke, 2006; Hsieh & Shannon, 2005). When the purpose of an analysis is to judge the quality, importance, value or effectiveness of a topic within the dataset, then evaluative analysis is a useful method (Patton, 2015). Evaluative analysis can be performed in different ways, like interviews, focus groups, surveys, and statistical analysis. This method is commonly used when the goal of the research is to make informed decisions based on evaluation outcomes, like program evaluation, policy analysis, and various forms of business and educational research. Evaluation is the process of generating information about the operations and impact of implemented programs or policies (Hennigan, Flay, & Haag, 1979). To conclude a program's effectiveness, evaluative research uses scientific methods in program evaluation (Hennigan, Flay, & Haag, 1979). Type-building qualitative text analysis is a methodological approach that involves categorizing and interpreting qualitative textual data by constructing types or categories to identify patterns, themes, and relationships within the data (Kuckartz, Three basic methods of Qualitative text analysis, 2014). To match the purpose of the research, the research method must allow a deeper assessment of the value and impact of the content, instead of categorizing or interpreting it, where qualitative content analysis focuses on identifying patterns, or type-building analysis constructs typologies, evaluative analysis provides a better understanding of the effectiveness and implications of a text (Kuckartz, Qualitative Text analysis: a Systematic approach., 2019; Hsieh & Shannon, 2005). In this research, we will conduct semi-structured interviews according to the evaluative analysis. We will use the evaluative analysis to code the interviews in a deductive way, so we will find out what the most important chapters in the different interviews are. Those themes will be used to formulate the assessment format that will be used for the rest of the research. The evaluative analysis will be executed by analyzing semi-structured assessments by experts. The assessments are semi-structured because they are carried out using an assessment format. We will implement evaluative analysis on the base of the assessment format. By using this standard document, it will give the experts who assess the documents a defined format. Using a structured abstract format helps to state conclusions clearly and concisely (West, Marsden, Humphreys, & Darke, 2018). All the assessments have the same base structure because of this format. Based on the aforementioned information, and the best use for each method, the evaluative analysis is the best qualitative analysis to execute this research.

3.1.2 Gioia Method for coding semi-structured interviews

In this research, we will be using an inductive method only to attempt to create a framework for coding the interviews. An inductive approach involves generating theories or patterns from the data itself (Dillon, 2012; Ketokivi & Choi, 2014; Azungah, 2018). A deductive approach starts with existing theories or hypotheses, and the data is analyzed to confirm or refute these preconceived notions (Azungah, 2018). An inductive method will be used because the input from experts will be used to organically formulate the assessment format in the best-customized way possible for this research. We will build an understanding of the data to develop theories by using this method. The interviews in this research will be held to identify what variables are important to assess the LLM-generated texts thoroughly. To identify and derive the important variables from the interviews, the interviews need to be coded. Coding allows the research team to systematically categorize and interpret the unstructured collected data, making it easier to identify variables (Côté, Salmela, Baria, & Russell, 1993). The variables for this research will be discovered through an informal inductive coding approach. An informal coding approach is more a flexible and exploratory approach where codes emerge organically from the data without predefined categories (Basit, 2010). Formal coding involves a structured, systematic approach to data analysis (Braun & Clarke, 2006). The coding approach will be informal because the interviews need to provide the key variables to make the assessment format as suitable for this research as possible. By analyzing the data this way, codes are developed and will emerge from the data during the analysis. The method that we will use to code the interviews will be the Gioia method (Magnani & Gioia, 2023). The Gioia Method is a systematic approach to qualitative research analysis that enhances the rigor and transparency of data interpretation (Magnani & Gioia, 2023). The Gioia method is a systematic approach to qualitative data coding and analysis that emphasizes inductive reasoning to build grounded theory from raw data. This method is suitable for our research because this method is particularly useful for inductively building theories (Magnani & Gioia, 2023). The Gioia method consists of three stages, first-order coding, second-order themes, and aggregate dimensions (Magnani & Gioia, 2023). The first step involves identifying first-order concepts using the participants' own language (Magnani & Gioia, 2023). The second step is organizing these concepts into second-order themes that reflect theoretical insights (Magnani & Gioia, 2023). The third stage involves the distilling of these themes into overarching aggregate dimensions (Magnani & Gioia, 2023). This approach provides a transparent and rigorous link between the qualitative data collected and the theoretical framework developed, enhancing the credibility and trustworthiness of the research findings (Magnani & Gioia, 2023). The variables for the assessment format are the codes that are mentioned as aggregate dimensions that are visualized in the coding scheme in Appendix D. The key variables that derive from the interviews and coding scheme are relevance, completeness, factual correctness, text clarity and structure, and time spent.

3.1.3 Evaluative analysis for formulating Assessment Format

The evaluative text analysis will be used to hold semi-structured interviews and to formulate the assessment format (Appendix A). The assessment format will be used by the experts to assess the LLM generated texts in a delineated format, making the eventual assessments based on equivalent basis (Sadler, 1989). The objectives of the assessments that were derived from the coding scheme as stated in Chapter 3.1.2, will be the following: Relevancy, Completeness, Factual correctness, Text clarity, Text structure, Overall judgment, Time spent for the assessment, and time required to make the outcomes applicable in an actual M&A process. These variables are crucial to determine how applicable LLMs really are in the internal and external analysis of an M&A process and will be elaborated in Chapter 3.2.1. and its subchapters. They are crucial because the variables are important factors at Taurus when the writing process commences. When an IM is reviewed by an employee at Taurus, the variables that were mentioned in this paragraph are taken into account.

3.2 Research Design

Several interviews with experts from Taurus will be held, for the qualitative analysis. The interviews will be held before generating the IMs by the different LLMs. They will be held beforehand because the information gained from the interviews will be applied to the formulation of the standard prompt. These interviews will provide a broad view of the M&A process. An interview with an AI expert will be held to determine what possible applications of LLMs could be useful for Taurus. The answers in the interviews will provide information for the qualitative analysis and the KPIs mentioned below. The interview guides are attached in Appendix B and C. The interviews with different experts will give a broad picture of the application of LLMs in M&A. The interviews will provide information on how to formulate the standard prompt and the variables that the LLM-generated texts need to comply with. On the one hand, the interviews with Taurus experts will show what the conditions are of a wellestablished IM. On the other hand, the AI expert will provide information on how to efficiently formulate a prompt. An overview of the research methodology is shown in Figure 2. The experiment part will be shown in more detail in Figure 3. A prompt can be more efficient if it generates output more quickly. The prompt design will be elaborated on in Chapter 3.3.

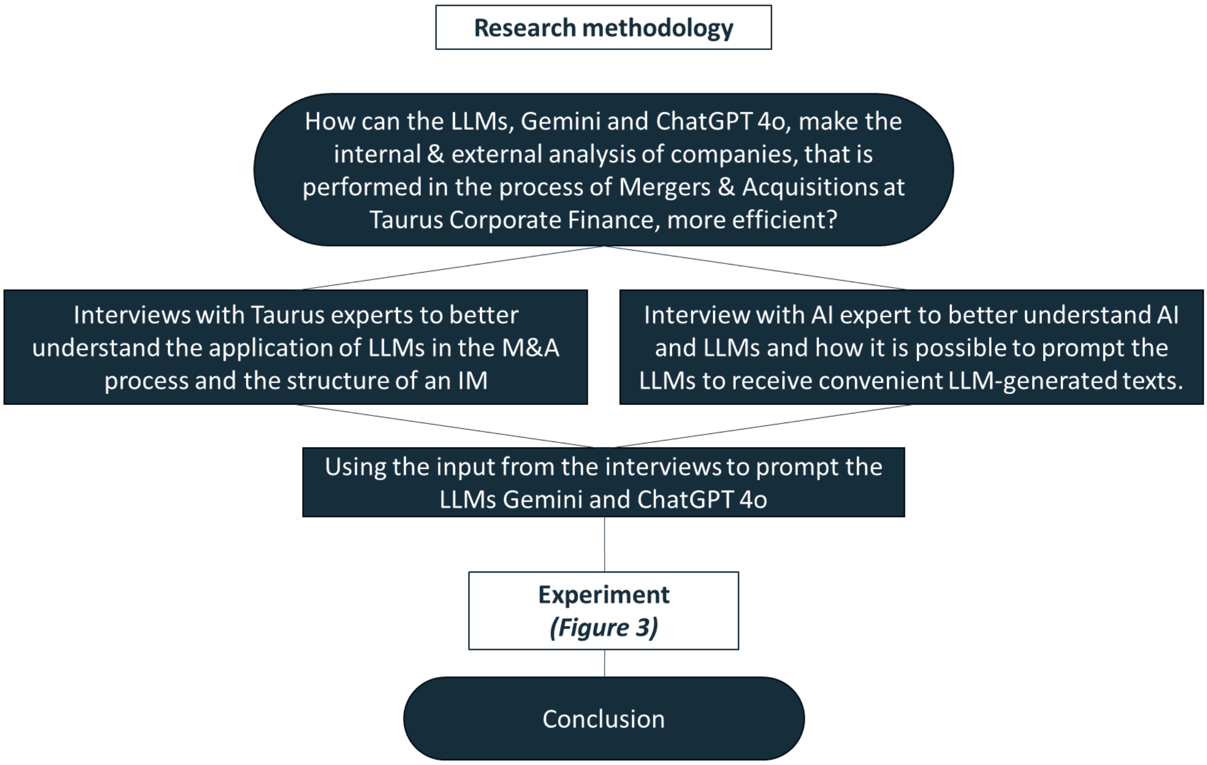

Figure 2. *Research methodology shows how we will try to answer the research question. The research question will be answered based on interviews with Taurus experts and an AI and IT specialist, and by executing the experiment of generating and assessing IMs, giving insight in how LLMs can make the internal and external analysis of companies more efficient.*

There are a lot of LLMs that currently exist, and the technology is developing rapidly. The number of active LLMs is increasing. For this research, existing LLMs will be used. The LLMs that will be used are Gemini and ChatGPT-4o. There is a possibility to create a custom-made LLM by yourself. (Topsakal & Akıncı, 2023) However, for this research, we use existing LLMs, instead of creating an LLM for this research.

In this research, only IMs will be used that are already created by Taurus, as described in Chapter 1.3. It would be difficult to gather IMs from other M&A corporate finance companies because the information shown in the IMs is sensitive and confidential. IMs will only be sent to parties that sign a Non-Disclosure Agreement (NDA). Therefore, it would be hard to use an external IM. Especially when these IMs would be used in prompts in LLMs. So, it will not be possible to use external IMs. The sensitive information in the IMs from Taurus will nevertheless be used for this research, but will not be disclosed in the thesis. Only the results of the assessments will be shown.

This research prompts ChatGPT and Gemini to create texts. Prompts in LLMs are instructions given to enforce rules, automate processes, and ensure specific qualities of generated output (White, et al., 2023). The generated texts are compared to texts written by Taurus' experts, to find out if the LLMs give useful analysis of companies. The texts will be compared by the Taurus experts. During this process, we will find out what the best possible prompt would be to generate the most useful text. Through trial and error, we will investigate what the best prompt would be for the internal and external analysis. A first indication of what prompt would generate the best text for the final assessment is possible to be executed by the research team. The research team will create a standard prompt that will be used to generate the texts. The research team will compile the LLM-generated texts. The Taurus experts will not help with creating the LLM-generated texts. The Taurus experts will only assess the final LLM-generated texts to conclude which LLM would be the better model for the internal and external analysis of companies, and if the LLM-generated texts are useful at all.

When an LLM is set up with the standard prompt, the text of the IM will be divided in smaller sections. Every prompt for this smaller section is unique in the analysis. To let the LLMs make an accurate analysis, it is important to determine the basic elements of a company to formulate in the prompt. Examples of those elements are the business model, the products that it sells, geographical presence, the sector that the company operates in, the size of the company, etc. This information is not sensitive and is publicly available. When this information is determined, it will be possible to make the most accurate prompt for every company. Sensitive information about customers from Taurus will not be uploaded to the LLMs. Sensitive information will only be used for this research to compare with LLM-generated texts. The LLMs will generate a text when it is prompted. A small change in a prompt can lead to a completely different outcome of the answer that is given by the LLM (Ye, et al., 2023). Through trial and error, we will find out in this research whether it is the case that small changes lead to different outcomes. The LLM-generated texts will be assessed by various Taurus experts according to the assessment format, as mentioned in 3.1.2 (Appendix A). The assessment formats are the same for every assessment. Therefore, an equal comparison can be made between the different LLMs and IMs.

3.2.1 KPIs

Various variables will be tested in this research. Each variable will be substantiated for why and how this variable will be measured. All the variables that form the basis of the assessment format are derived from the coding scheme of the interviews. The assessment format contains several quantitative questions. The answers that those questions will provide, will be used to create graphs to visualize the outcomes. Graphs can improve viewers' interpretations by minimizing inferential processes and maximizing pattern association processes (Shah, Mayer, & Hegarty, 1999). The understanding of the difference between scores in IMs is important to conclude why certain IMs receive better scores compared to other IMs.

3.2.1.1 Relevancy

The variable relevancy has been taken into account in the assessment format. What is meant by a relevant text in this research, is that the text contains company specific information about a certain topic, instead of general information about a certain topic in the analysis (Lin, et al., 2024). A text can more easily be applied in an actual IM, when an LLM-generated text is relevant. The relevancy of a text will be assessed in several questions in the assessment. The questions will provide the research team with a clear understanding of the relevancy. The following questions are mentioned in the assessment:

How relevant is the text that is written in this analysis?

Can you assess the relevance of this text on a scale from 1 to 10?

← Bad					*Good →*				

3.2.1.2 Completeness

The experts at Taurus, who actually wrote the original IMs, will be able to determine whether LLM-generated texts are complete, for that specific IM. When key information about a company has not been accounted for, then the quality of the analysis would be insufficient

(Veen, et al., 2023). This KPI is important to determine whether LLMs leave out information. The question in the assessment that will be used to conclude this KPI, is the following:

How complete is the text?

3.2.1.3 Factual correctness

The KPI of factual correctness can also be assessed by the experts who wrote the original IM. Factual correctness is related to completeness of the text. However, where completeness assesses whether all information is included in the text, the KPI factual correctness will determine if the written information in the text is correct (Chen, et al., 2023). Incorrect information in an IM will have a negative effect on the final stage of an M&A process of a company, when this incorrectness is discovered in the due diligence process (Courteau, 2020). This could result in a negative revision of the price of the company, or it could mean that the process will not be continued, depending on how the buyer assesses the incorrect information. The question that will be used in the assessment to conclude this KPI is the following:

Is the text factual correct, compared to the original text that was used in the actual IM?

3.2.1.4 Text clarity and Text structure

The KPI's Text clarity and Text structure are not depending specifically on the companies. These KPI's represent the ability of the LLMs for which they master the language that they wrote and these KPI's are essential for assessing both usability and effectiveness. Text clarity ensures that the generated content is easy to understand, minimizes ambiguity, and conveys the intended message accurately (Meyer, 2003). Text structure refers to the logical and coherent organization of information, which is crucial for maintaining the reader's engagement and ensuring that complex ideas are presented in a way that facilitates understanding (Meyer, 2003). The questions in the assessment that pertain to text clarity and text structure are the following:

Can you assess the clarity of this text on a scale from 1 to 10?

← *Bad*					*Good* →				

Can you assess the structure of this text on a scale from 1 to 10?

← *Bad*					*Good* →				

3.2.1.5 Overall judgment

The overall score for the LLM-generated text will be assessed by a quantitative question. This KPI is the most important to conclude which LLM performs better, the way that the LLMs is used in this research. Comparative judgment is a strong method for assessing text quality, as it minimizes differences between assessors and focuses on argumentation and organization in a holistic manner (Lesterhuis, Bouwer, Daal, Donche, & Maeyer, 2022). Therefore, it is important to conclude why a certain IM scores higher compared to the other IMs. When we understand this, we can use that information to customize a prompt in such a way that it generates a better IM for that specific company. The overall variable is important to conclude which LLM would perform its task better compared to the other when we account for all the variables. The quantitative question that belongs to the KPI overall judgment is the following:

What is your overall judgment of the text on a scale from 1 to 10?

← *Bad*					*Good* →				

3.2.1.6 Efficiency

An important KPI for this research is efficiency, where the amount of time the use of LLMs would save Taurus consultants during their work. The reason that this variable comes after the overall judgment is practical. The assessors can only decide how long the assessment took on the end of an assessment. We will investigate how much time Taurus consultants would spend on writing the internal and external analysis of a company. The answer will be given in the interviews that will be held with the experts from Taurus. The time that is compared is purely the time that is spent to write the internal and external analysis. The time that it takes to create the visuals will not be considered, since this does not fit in the research design, as described in Chapter 1.2. By interviewing several experts at Taurus, it is possible to estimate the average time it would take for an expert to write an internal and external analysis. The time that an LLM would require to write a text will be considered to be minimal. LLMs create texts very rapidly, and pure writing is always faster when it is done by LLMs (Witteveen & Andrews, 2019). However, LLM-generated texts need to be verified and possibly corrected by experts (Liu, et al., 2023). LLM-generated texts could contain mistakes or inaccuracies because an LLM does not always use sources to generate a text. LLMs can generate text with citations, improving their factual correctness and verifiability, but current systems have room for improvement (Gao, Yen, Yu, & Chen, 2023). Implementing LLMs in the internal and external analysis process would change the work of Taurus' consultants. The activities would change from self-analyzing and writing the texts, to verifying, analyzing, and correcting the LLM-generated text. The consultants still need to analyze the company, otherwise they would not know whether the written texts are accurate and complete. This verification and possible correction of the LLM-generated analysis does require time and experience. The time required for analyzing the company would not change when this process changes, because it needs to be done in both scenarios. The time to verify and correct the LLM-generated texts will be compared to the average time of actually writing the analysis itself. This research will give insight into the (potential) time reduction when LLMs are implemented in the process, and if there is a difference. The questions that belong to the KPI efficiency in the assessment are the following:

How much time did you spend assessing this text?

How much time would you expect to spend to make this text applicable for an IM?

3.2.2 Experiment

To conduct accurate research about the application of LLMs, we will compare the LLM-generated text results from 2 different LLMs with 5 different IMs that Taurus' experts have written in English. The research will be conducted with recent IMs, because these IMs have a higher quality, and are more relevant when it comes to external analysis. They are more relevant because the external environment is changing continuously, and older IMs could have completely different conditions. For example, an IM that was written in 2019 will not provide any information about the Covid-19 pandemic, which can be very relevant for some companies. The companies of the used IMs all have their specific sector in which they operate. The sectors of those companies are very diverse. The sectors are: E-commerce company in the clothing industry, Recycling company, Cold and freezer warehouse, E-commerce of electrical wholesale, and a Business service provider. Because the sectors are so diverse, the research itself will be more complete and more diverse.

The 5 IMs are the base texts and will be compared to the outcomes of the 2 different LLMs. In the end, this will result in 10 different comparisons with the help of the evaluative analysis by the experts at Taurus. Because the research must remain manageable, the number of IMs is relatively small. To generate and assess an IM is time-consuming, therefore a profound qualitative analysis on a smaller number of IMs is better. Most IMs are assessed by more than one expert, to reduce a bias from single experts. This way we create a substantiated result if LLMs could be applied to internal and external analysis of the M&A process of a corporate finance company through qualitative analysis. This research will answer if LLMs can assist in this process, and if it can accelerate the process. Furthermore, this research will give an indication for which LLM is the most capable for the application in the M&A process of a corporate finance company. The experiment will be conducted by asking the LLMs to generate an internal and external analysis of the sector in which a specific company operates in, through varying prompts. By conducting the experiment this way, it will be possible to determine what LLM would generate the most applicable texts. To create a feedback loop in the experiment, the results of the assessments will we evaluated to make possible changes in the prompt, improving the LLM-generated texts. An overview of all the steps in the experiment can be seen in Figure 3. The most applicable texts are texts that need relatively few corrections to make them relevant for an IM.

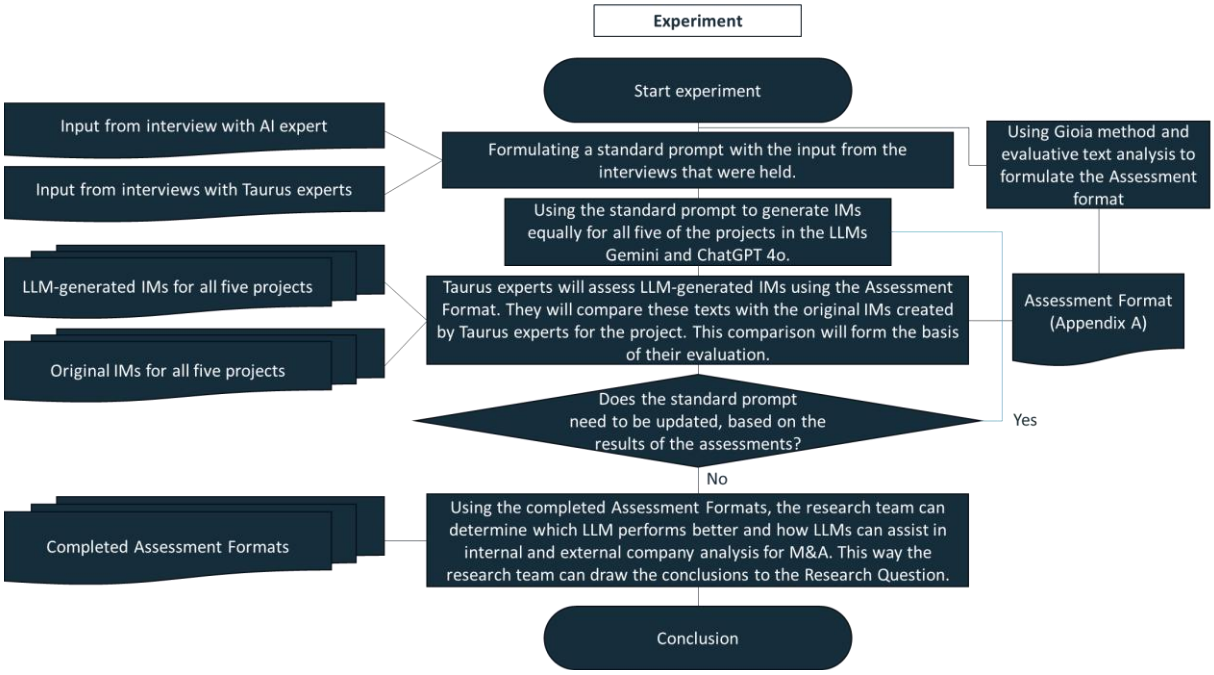

Figure 3. *The way the research team will execute the experiment. The LLMs ChatGPT 4o and Gemini will be prompted to generate a text that could be used in an IM. This output will be assessed by Taurus experts, while they are also compared to existing IMs. Showing how LLMs can make the writing process of the internal and external analysis more efficient, and show which LLM works better in which category.*

3.3 Prompt

To let an LLM generate a text, it is necessary to put a prompt into the model. This prompt will result in a conversation between the user and the LLM. In this chapter, we will explain what the specifications are, what the segments of the prompt are, and why they were implemented in the standard prompt. The standard prompt was developed through trial and error, to determine what prompt would generate the most applicable results. Based on the interview with the AI expert, we now know that there is not one right way to prompt in an LLM. *"Look, the best way to prompt is to ask ChatGPT itself. And then it will give you different possibilities for getting the best result. It involves a lot of trial and error. So, before you figure out what you get, you will first want to develop a prompt yourself. A prompt is nothing more than just an input query."* (Interview D) But, there are several things that you can take into account to make the LLM give more applicable results. *"It is a very short, clear, efficient formulation of your goal."* For every prompt that you make, you first have to make sure that you know what the purpose is of that particular prompt. *"The best way to figure out what the best way to prompt* *is, is to use trial and error."* While generating IMs for the assessment, we have used various prompts. By using those various prompts, it was possible to determine what were crucial parts of the prompt. One of the most important parts is the sentence: *"Please keep in mind that this is a real company!"*. When this part was not added to the prompt, the results would be a lot worse than when this part was added. The AI expert helped formulate a base prompt with a few suggestions. This prompt was updated gradually by adding parts each time to receive more useful texts when this was deemed necessary. Eventually, this resulted in a standard prompt that is used at the beginning of a conversation with the LLM to set up the LLM correctly. The structure of the prompt will be explained in this chapter below. The most crucial parts of the prompt are written in black. The less important parts are written in red. When the red parts of the prompt are removed, the generated texts are still useful and are not very different from the results of the standard prompt. However, when you remove all the red parts, the results would not meet the standard qualifications. The combination of the different aspects and the specific guidelines that the generated text must comply with according to the prompt, is the key to success when using the LLMs.

3.3.1 Standard Prompt

The standard prompt that was used to generate the LLM-generated texts:

You are a corporate finance professional. *Your goal is to highlight the strengths and growth potential of a company to attract strategic buyers for a potential acquisition.*

The target audience for the text is strategic buyers. *The text should be professional but also enthusiastic to generate interest and encourage them to contact you for further discussions.*

After each response, ask the user a question to enrich the text for a potential buyer based on your knowledge of the market and company.

When writing the text, you must adhere to the following principles. You will be penalized if you do not follow these principles.

- *Your tone is always kind, empathetic, motivating, and human.*
- *Use a pyramid writing style (most important information first)*
- *Use B1-level English combined with regular spoken language*
- *Avoid unnecessary helping verbs*
- *Maximum 15 words per sentence*
- *No abbreviations*
- *No jargon*

- *Active writing style*
- *Write in the third person singular*

Describe the company [company name]. Emphasize the unique value proposition, competitive advantages, financial performance, and growth opportunities. Conclude with a clear call-to-action that encourages potential buyers to contact you for more information.

Please keep in mind that this is a real company!

Use relevant sources, like their website, CBS, Eurostat and/or branch organizations.

3.3.2 Prompt Structure

The prompt was written by the research team, in combination with two employees at Taurus who are responsible for AI implementation at Taurus, and the AI expert who was interviewed for this research.

Once the LLM is prompted with this standard prompt, it is not necessary to use this standard prompt over and over again. The conversation with the LLM will be using the guidelines from your standard prompt for the rest of the conversation. We also found out that it is very helpful for the end result to split the IM in different chapters. This way, the LLM will generate shorter texts that you can use for a single paragraph in the IM. By dividing the IM in subchapters, the LLM is capable of generating a more detailed and complete IM. When the LLM is prompted to generate an entire IM in a single prompt, the results are not sufficient to use in practice. The prompt is divided into different segments each representing the conditions the LLM-generated output must comply with.

3.3.2.1 Role of the model

You are a corporate finance professional. *Your goal is to highlight the strengths and growth potential of a company to attract strategic buyers for a potential acquisition.*

This segment is used in the prompt because this sets up the LLM that it could account for from what perspective the text needs to be written. This will trigger the LLM to write from a corporate finance professional and sales point of view.

3.3.2.2 Introduction to the target audience

The target audience for the text is strategic buyers. *The text should be professional but also enthusiastic to generate interest and encourage them to contact you for further discussions.*

Segment 2 is used in the prompt because it gives the LLM a clear instruction for what the target audience of the eventual text would be. Therefore, the LLM can account for this target audience for the rest of the conversation.

3.3.2.3 Feedback loop

After each response, ask the user a question to enrich the text for a potential buyer based on your knowledge of the market and company.

This part of the prompt creates a feedback loop and is included as it helps to enrich the final text. The employee at Taurus can provide the LLM with additional information that the LLM has not accounted for, or could not know in the original generated text. This part of the prompt is red because in most cases the Taurus employee cannot give this information to the LLM. The information that selling companies provide to Taurus is classified. Therefore, it is undesirable to present the LLMs with this information because the model will use this information for machine learning. Information that is uploaded in the model is not secure. However, when you use a version of the LLMs where the information you put in would not be used for machine learning, then this segment of the prompt could prove valuable. Using LLMs that would not use your information for machine learning, was not possible for this research. In ChatGPT, you can create your own GPT that does not use your information for machine learning. This option is not available in Gemini. Therefore, if we used this secured version of ChatGPT, the provided information in both models would differ. This would create an unbalanced research. In summary, Segment 3 of the prompt is very useful, but only when you use a secured model that does not use your information for machine learning.

3.3.2.4 Minimum standards of the output

When writing the text, you must adhere to the following principles. You will be penalized if you do not follow these principles.

- *Your tone is always kind, empathetic, motivating, and human.*
- *Use a pyramid writing style (most important information first)*
- *Use B1-level English combined with regular spoken language*
- *Avoid unnecessary helping verbs*
- *Maximum 15 words per sentence*
- *No abbreviations*
- *No jargon*
- *Active writing style*
- *Write in the third person singular*

This segment is used in the prompt because this is the house style of Taurus when writing any text at Taurus. So this house style is used for IMs, Letters, emails, teasers, etcetera. This segment helps the LLM to write in the Taurus style, which makes it easier to use the LLM-generated text. The red sentence, *"You will be penalized if you do not follow these principles"* was added because the AI expert that we interviewed told us that LLMs would respond more accurately to the prompt when the user threatens to punish the LLM when it would not execute the prompt correctly. However, after testing both variations, we concluded that the difference between both prompts was limited. Therefore, adding a threat to punish the LLM will not lead to completely different results.

3.3.2.5 Task instruction for the model

Describe the company [company name]. Emphasize the unique value proposition, competitive advantages, financial performance, and growth opportunities. Conclude with a clear call-to-action that encourages potential buyers to contact you for more information.

Segment 5 is added to the prompt because this is the task that you present the LLM with. Without this segment, the prompt would not have a clear instruction for what you would want the model to do.

3.3.2.6 Directive for the model

Please keep in mind that this is a real company!

This segment was added in the prompt because, through trial and error, we found out that the LLMs assumed a company to be fake when we mentioned it. By adding this segment in the prompt, the LLM would present far more reliable information, than when this segment was not added. Therefore, this is one of the most important segments to increase the reliability of the LLM-generated texts.

3.3.2.7 Guideline to take into account

Use relevant sources, like their website, CBS, Eurostat and/or branch organizations.

Segment 7 was added to the prompt to help enrich the LLM-generated text with actual information that relates to the company that the LLM needs to write about. For the external analysis of the company, this segment would trigger the LLM to use those relevant sources.

3.4 Strength of the experiment

The assessments by Taurus experts need to be validated to ensure that the texts are assessed in the most substantiated way possible. This research will use the Evaluation analysis (Chapter 3.1.3) to judge how accurate the LLM-generated texts are. For this research, several experts from Taurus will judge the LLM-generated texts. The experts that will judge the generated texts, were part of the deal team of the original project or is the employee that usually reads every document within the framework of the four-eyes principle at Taurus. An IM at Taurus is always assessed internally by Assessor 1. He will also assess every LLM-generated text. He is the founder of Taurus and is the most experienced manager within Taurus in the field of M&A. Therefore, the assessments will be executed by experts that either have knowledge of the original IM, or have the experience to assess the LLM-generated texts. Because several experts judge all the LLM-generated texts, they would automatically judge what LLM provides the experts with the most applicable results for the internal and external analysis. This makes it possible to conclude which LLM would be better for the application in the field of M&A.

The research will be executed as described previously in this chapter. Only one LLM-generated text was assessed by one Taurus expert. The rest of the LLM-generated texts were assessed by two or more Taurus experts. Therefore, almost every LLM-generated texts complies with the four-eyes principle to give a more substantiated conclusion. When the results of the assessments show similar findings about which LLMs would generate better texts and how the LLM-generated texts can be used in the internal and external analysis, then we can conclude which LLM would be the better model.

4. Results

This chapter presents the findings obtained from the interviews and the LLM-generated texts that Taurus experts assessed. In this chapter, we will see what prompt was used, and how this prompt is constructed. Furthermore, we will dive deeper into the results and findings of those assessments.

4.1 Results Interviews and Assessments

This subchapter presents the results from the interviews and the assessments, and the similar findings in both sections of this research. The interviews were held with three different experts from Taurus, and one AI expert, as mentioned in Chapter 1. The assessments were conducted at Taurus by in total 6 different experts. The LLM-generated texts were sent to the experts, along with the text of the original IM, an empty assessment format, and a short instruction for the assessments. The assessments all present a similar result for the IMs, except one assessed by multiple experts. The experts agree on almost every question in the assessment format. In general, Gemini generates a more useful output for the purpose of the texts at Taurus.

4.1.1 Quality of the Prompt

When we look at the qualitative questions in the Assessments, then we see something remarkable. IMs that were generated later, chronologically speaking, are considered to be of better quality. Those IMs received a higher score on the quantitative scales, compared to the IMs that were generated earlier. The feedback that was given on the IMs that were generated earlier, was applied on the conversations with the LLMs. As stated in Chapter 3.3.2. Once the LLM is prompted with this standard prompt, it is not necessary to use this standard prompt over and over again. Based on the received feedback on the first IMs, it was easier to formulate the consecutive prompts to generate the entire IM of that Project. This resulted in better LLM-generated texts. The chronological order for generating the IMs was FREEZE, DIAMOND, SCRAP, WIRE, and OMNIA.

4.1.2 Quality of the Source

What is also important for the assessment results, is the quality of the source that the LLMs use to generate the IMs. The most important source for the LLMs to generate the texts are the websites of the companies from the IMs. Companies with a better website and more public information generate a better quality IM. Projects SCRAP and WIRE both have good quality websites. The companies are a lot bigger than the other companies that were assessed. Project FREEZE on the other hand, has a very poor website. What we see in the assessments of FREEZE, is that the LLM-generated texts were worse than other assessments. This means that the quality of the website of companies is a major requirement, to receive a better LLM output.

4.1.3 Factual Incorrectness

We can see an important similarity when we look at the results from the interviews and the assessments. Factual incorrectness is one of the most important threats when using LLMs. This observation was stated by the AI expert, and by all the Taurus experts that were interviewed. "*What people often do not realize is that it does not truly understand language". "You need to be aware of that". "So, there is still a challenge in making sure you verify it". "One of the main dangers is that the user might become complacent, and you may not have a clear way to verify what is true"* (Interview D). These citations were drawn from the interview with the AI expert. He mentions multiple times that you should always review and possibly adjust LLM-generated texts. The Taurus experts that were interviewed all stated something about the threat of misinformation. *"You still need to verify those companies because it is risky to blindly trust a suggestion"* (Interview C). *"Checking whether it is misinformation is important"* (Interview A). *"If you use AI without that knowledge, you cannot validate whether what comes out is accurate or logical"* (Interview B). These are only a few examples of every interview. However, the threat of misinformation is mentioned several times. This indicates that the experts at Taurus are very well aware that you should always review the output of the LLMs.

As we have stated before in Chapter 3.2.1.6. the time required for analyzing the company would not change, because this needs to be done in both scenarios. This complies with the following statement in the interview with Taurus expert Interviewee 3. *"You can only use AI effectively when you have enough background knowledge to assess whether the conclusion it draws is close to the truth. If you use AI without that knowledge, you cannot validate whether what comes out is accurate or logical."* (Interview B). Factual inaccuracies are a very important limitation of LLMs and therefore all output needs to be reviewed.

4.1.4 Relevancy of LLM-generated texts

The LLM-generated texts are very relevant to use for the analysis of companies. This can be seen by the following quotes from the assessments. *"I see little text that does not belong to the case study. There is relatively much overlap through the topics though. But so is our own*

IM" (Appendix P). *"Partially relevant, especially some summarized sentences could very well be used in an IM"* (Appendix N). Because the output of LLMs is so relevant, it could help consultants in their writing process in several ways. It could help the writing itself, but this shows that the LLMs could help consultants inspire them. The LLMs can offer a broad view of the company, that can help consultants with their analysis. This is stated in Appendix I, J, and M, *"The broad market analysis. Especially in this case where the seller has little market input". "Offering a broader view". "The combination of market intel and company information. It gives the impression of having a broad scope".* LLMs can generate texts about certain topics of a company, that a consultant had not thought about. This is shown in Appendix L. While generating the text for the analysis of Project DIAMOND, Gemini generated an analysis that had not been done in the original IM. Therefore, this inspiration by the LLMs could help consultants make analyses that they had not thought about in the first place.

4.1.5 LLMs cannot understand context

LLMs lack the ability to understand context while generating texts. This is a weak characteristic of LLMs. When the LLMs are not prompted correctly, they will generate extended texts. This is another reason why LLM-generated texts cannot be applied unconditionally and need to be reviewed at all times. The LLM will, for example, exaggerate a very minor thing about a company, instead of more important topics. This was mentioned in the assessment in Appendix M, *"Sometimes overly simple texts and a bit too optimistic. Strong points of the company are being exaggerated".* This connects to what the AI expert stated in the interview, (Interview D), that LLMs could not understand language. *"A language model is just a very good predictor and does not always understand the context". "It is simply a data magician that you can teach. But understanding context is very difficult for it". "It is a language magician. What people often do not realize is that it does not truly understand language".* This shows that LLMs could lay emphasis on wrong and less important topics in analysis, making the analysis less valuable.

4.1.6 Quantitative Part Assessments

The results of the quantitative analysis show a couple of standouts. Gemini scores better in terms of clarity, relevance and overall score, the text structure is similarly assessed. The results of the analysis are shown in Figure 4 and Table 1.

Gemini scores better in terms of clarity on a difference of 1,3 points average out of 10. Gemini scores 7,3 points compared to 6,0 points for ChatGPT out of 10. Based on the results for text clarity, we can conclude that Gemini generates clearer texts than ChatGPT.

The relevancy of the text is assessed better for Gemini than ChatGPT. Gemini scores 7,1 points on average for relevancy compared to ChatGPT which scores 5,0 points on average out of 10. This is a significant difference of 2,1 points. We can state that Gemini scores significantly better on text relevancy than ChatGPT.

The text structure of the LLM-generated texts are assessed similarly for both LLMs. The average difference between the two LLMs is 0,1 point out of 10. Gemini scores 6,4 points compared to 6,3 points from ChatGPT on average. Even though Gemini scores a bit higher than ChatGPT, we cannot conclude that Gemini scores better on text structure compared to ChatGPT.

The overall score of both LLMs show a clear and wide difference between the two LLMs. Gemini scores higher than ChatGPT by 2,0 points difference on average. Gemini has an overall score of 6,6 compared to an overall score of 4,6 for ChatGPT. Based on this significant difference in scores, we can conclude that Gemini is better than ChatGPT overall.

Based on the results of the analysis we can conclude that Gemini is the better LLM for the application that was explored in this research.

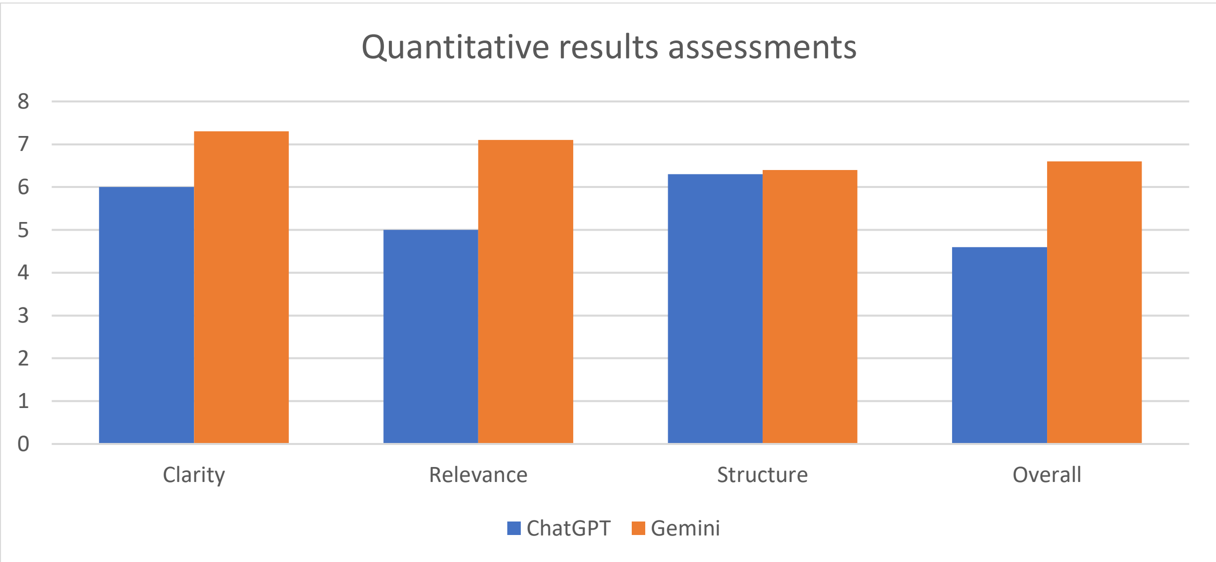

Figure 4. *Quantitative assessments on the average scores of the LLM-generated texts. Gemini scores higher on all assessments. There is clearly a higher score for Gemini, however for text*

structure the difference is 0,1 point on average. The data to create this graph can be seen in Table 1.

	ChatGPT				Gemini			
Assessor	**Clarity**	**Relevance**	**Structure**	**Overall**	**Clarity**	**Relevance**	**Structure**	**Overall**
1	2	2	5	3	3	3	5	3
1	3	2	6	1	9	7	7	8
2	6	6	7	5	8	9	8	8
3	9	4	6	5	8	8	3	6
1	9	7	9	3	9	9	9	8
4	2	3	1	3	4	5	2	4
1	8	8	9	8	9	9	9	8
5	7	6	6	6	7	6	7	6
2	7	6	7	6	8	8	7	8
6	7	6	7	6	8	7	7	7
Average scores	6,0	5,0	6,3	4,6	7,3	7,1	6,4	6,6

Table 1. *The average scores of the assessment of all the different quantitative questions. The results show that Gemini scores higher than ChatGPT on every question. This data was used to create Figure 4.*

5. Conclusion

The conclusion chapter synthesizes the key findings of this qualitative research. The expert interviews and assessments provide a substantiated answer for the sub-research question 4 *“Which LLM is the most suitable to improve the process of generating analytical reports of companies, and to improve the efficiency of the analysis for Taurus?”* and the main research question *“How can the LLMs Gemini and ChatGPT 4o make the internal & external analysis of companies, that are performed in the process of Mergers & Acquisitions at Taurus Corporate Finance, more efficient?”*. The experts who assessed the documents all came to a clear conclusion regarding the LLM-generated IMs. In general, Gemini will provide better textual answers for the prompts that were given in the LLMs. This is in line with the hypothesis that was stated in the interview with the AI expert, (Interview D). *“You also have models that are better with languages”. “Gemini has made a huge leap in that”.* The results of the assessments are consistent with the predictions of the AI expert.

5.1 Public Information

The quality of the company’s website is crucial for the usability of LLMs. This was mentioned in Chapter 4.1.2. Project FREEZE is a company that has an insufficient website, especially compared to the other Projects that were assessed. The analysis of the assessments points out that Project FREEZE is rated much worse, compared to the other projects, both for ChatGPT and for Gemini. On the other hand, Project SCRAP and WIRE were rated one of the best in the assessments. The websites of those companies approved to be very sophisticated and extensive. Even though the LLMs could use limited sources, the generated text could be useful as an inspiration for writing the IM.

5.2 Misinformation

Factual incorrectness is the most important threat of using LLMs. As previously written in Chapter 4.1.3, this was stated in every interview that was conducted. This is also shown in the assessments. Only in two assessments, there is nothing mentioned about any inaccuracies or misinformation. It is mentioned in every other assessment that the text is incorrect at some points. Therefore, as we have mentioned in Chapter 3.2.1.6, the time required to analyze the company would remain the same. Otherwise, you would not know where you need to correct the LLM. The LLM-generated output always needs to be reviewed, since you cannot unconditionally rely on the outcome of the LLM.

5.3 Relevancy Conclusion

The assessments show that the LLM-generated texts are quite relevant. What relevant texts mean, is that the texts could very well be used in actual IMs. The only condition for this is that the company's website must contain enough information. If that is the case, then the LLMs often provide usable text for an IM. This is shown in the assessments, FREEZE is not very relevant, whereas SCRAP and WIRE are very relevant. FREEZE has a very poor website, and SCRAP and WIRE have better websites and more information publicly available since the companies are a lot bigger. Therefore, LLMs can help inspire consultants when writing an analysis. The output can also be used for their structured presentation and their business writing style. Even though some parts of the LLM-generated texts are incorrect, they are still presented in a structured way. This structure could be taken over when writing an actual analysis for an IM. LLMs can generate analyses that consultants have not thought about. So this points back to the conclusion stated previously in this paragraph, LLMs could inspire consultants when writing an analysis.

5.4 LLMs generate lengthy texts

A weak characteristic of LLMs is generating very lengthy texts. When the prompt is not set up right, the generated texts could become too extended. However, when the prompt is formulated to generate a shorter text, it can disregard important topics, and not scrap out the less important information. LLMs can sometimes exaggerate texts that are completely insignificant in a context because the models cannot understand languages. As stated in Chapter 4.1.3, you cannot unconditionally rely on the outcome of the LLM. Therefore, a consultant would always need to review the LLM-generated texts. The consultant needs to know the key information and the business model of the respective company, to be able to review and possibly adjust the LLM-generated texts. Because LLMs cannot understand languages, it is possible that it could lay emphasis more on minor subjects, making that part seem important, even though it is not. The most important aspect of a good IM is the length of the text. The text needs to give an objective analysis and touch on the key items of a business case. Often the LLMs generate a too broad text to comply with these aspects. When

LLMs are prompted to use fewer words, then the risk arises that important information is left out. Therefore, the LLMs are not reliable enough to copy directly into an IM.

5.5 Quantitative Conclusion

The quantitative results from the assessments show a couple of conclusions. Gemini receives a better result than ChatGPT on every part of the qualitative analysis, which is also shown in Figures 4 and Table 1 as mentioned in Chapter 4.1.6. There is a small difference in the average scores of Gemini and ChatGPT on the subject of Text structure. The difference here is 0,1 point on average higher for Gemini compared to ChatGPT out of 10. This difference is marginal, which means that we cannot clearly state that Gemini generates a more structured text, compared to ChatGPT. In terms of text clarity, Gemini scores 7,3 points compared to 6,0 points for ChatGPT out of 10 on average. We can conclude that Gemini generates much clearer texts than ChatGPT. The scores for the text relevance give a clear conclusion. Gemini receives 7,1 points for text relevance compared to 5,0 points for ChatGPT out of 10, so Gemini generates more relevant texts than ChatGPT.

The overall score of the assessments shows that Gemini scored significantly higher on average than ChatGPT. The total scores are 6,6 points for Gemini compared to 4,6 points for ChatGPT. The difference is 2,0 points out of a total average score of 10. This clearly shows that Gemini right now is a better LLM to use for writing the internal and external analysis for companies in the process of M&A at Taurus.

5.6 Overall Conclusion

To answer the main research question, we can conclude that LLMs are useful for writing an IM in an M&A process at Taurus. However, the generated texts need to be reviewed and where necessary adjusted. When Taurus decides to work with LLMs in the M&A process, this would significantly change their way of work. The employees will change from input tasks to output control. This was also mentioned in an interview with a Taurus expert in Interview C. *"You will see a shift where people who used to do input tasks will move to output control".* The way that the employees would do their jobs would change significantly. The advantage for the consultants is that their research in the company would not change because this needs to be done in both scenarios, as stated in Chapter 3.2.1.6. Otherwise, it is not possible to adjust the LLM-generated texts when necessary. The writing process however will change tremendously.

Employees would need to learn how to work with LLMs and how to prompt them, as mentioned by the AI expert. *"That is the challenge between innovation with AI. If I want to do something with it, I'll need to make time. I need to learn to drive. It is a huge turning point that you need to start taking lessons"* (Interview D). To implement LLMs in the field of work, will take time. The employees need to gain confidence in how the models work. This can only be accomplished by regularly working with the models. In the beginning, this will probably only be rewriting and summarizing texts. Eventually, this will grow in more confidence to let LLMs generate analysis texts that could be used in an IM.

6. Discussion

There are some limitations to this research. This research was executed at Taurus Corporate Finance in Deventer. The IMs that were generated by the LLMs were originally made by Taurus experts. Other corporate finance companies would have written other IMs in the first place. That is why the LLM-generated texts are more focused on the Taurus style of writing. Therefore, the findings in this research cannot be generalized for other corporate finance companies. However, despite this limitation, this study makes an important contribution to the literature because this research makes it easier for other corporate finance companies to investigate this matter for their own company. The research was conducted only at Taurus because the materials used to generate IMs contain sensitive and often classified company information. This information was originally shared with Taurus under the conditions of a Non-Disclosure Agreement. This would make it hard for this research to gather IMs from other corporate finance companies. In future research at other corporate finance companies, this would automatically be solved, when that research will use their own internal IMs. The opportunity for future research is to perform a likewise research on the application of LLMs in the internal and external analysis of companies in an M&A process at other corporate finance companies, to create a more substantiated conclusion for how LLMs could assist in the analysis part of an M&A process. Therefore, the results and conclusions of this research cannot be adopted directly by other corporate finance companies.

Another limitation of this research is the use of only 5 IMs, and the use of two different LLMs to generate the IMs. This was done because the research needed to be manageable. The assessments require a lot of additional time for the Taurus experts besides their regular job. Therefore, it would not be in the interest of Taurus to generate and assess more IMs. To compensate for the limited amount of IMs that were assessed, we used IMs from very different sectors to mitigate using fewer IMs. To get a better view of what LLM would be the best to use for writing IMs, it would be better to use more LLMs. In this research, we have only investigated two models. For future research, more assessments on different IMs on more LLMs would give a more complete answer to the research question compared to this research.

A third limitation of this research is the assessment by 6 different experts. In total 6 different experts assessed the LLM-generated texts. Most of the LLM-generated texts were assessed by two experts, one LLM-generated text was assessed by one expert, and three experts assessed one LLM-generated text. If every text had been assessed by 6 different experts, the research would have been more thorough. However, the assessment of the LLM-generated texts takes a lot of time for the employees at Taurus, as stated in the previous paragraph. When every LLM-generated text would have been assessed by the 6 different experts, the number of assessments would increase from 20 now, to 60. This would result in a more widely supported conclusion. The experts did not have to assess all the IMs, because most of the time the expert that assessed the LLM-generated IM, was part of the deal team that wrote the original IM. When every expert would assess every IM, they would have to prepare themselves better to assess the LLM-generated texts. Because the experts were part of the deal team that wrote the original IM, this was more efficient in the assessments. Therefore, this was done to shorten the time required to assess the LLM-generated texts. This creates an opportunity for future research, as additional studies could involve a broader panel of experts to assess the LLM-generated texts on the same variables that were used in this research. This will provide more substantiated results for how LLMs could assist in the M&A process.

The generating process of this research was partly changed during the research. When IMs were recreated and assessed, the prompt was changed to improve the outcome of the next LLM-generated IM. This was done to create the best possible outcomes for the replicated IMs. However, this biases the outcomes between the different IMs, as the input was has been enhanced. This was slight contamination was performed to generate the best possible IMs as an end result.

The results show that the LLM-generated output sometimes is factually incorrect, this correlates with the literature about the use of LLMs. The spread of misinformation through the misuse of generative AI tools is a real and present danger (Borger, et al., 2023). LLM-generated misinformation can be harder to detect than human-written misinformation with the same semantics, potentially causing more harm (Chen & Shu, Can LLM-Generated Misinformation Be Detected?., 2023). LLMs can both bring promising opportunities for combating misinformation and easily generate deceptive misinformation at scale (Chen & Shu, Combating misinformation in the age of LLMs: Opportunities and challenges, 2024). LLMs can generate incorrect information, known as the hallucination problem, which can pose societal threats and negatively affect in-context learning (Saha, Ganguly, Saha, & Mitra, 2023). The findings in this research correspond to the theoretical background.

This research provides valuable contributions for both practice and theory because this is the first research to empirically investigate the application of LLMs in the internal and external analysis of companies involved in an M&A process. This will fill a part of the research gap, where very little is known about the application of LLM in general, and especially in an M&Â process. This research provides a practical contribution for employees at Taurus, as they can use this research to determine how they can use the LLMs to generate more useful texts for their practices. Readers outside of Taurus could also benefit from this research, as they could use this research to better understand how to prompt an LLM on their own.

Bibliography

Alberts, I., Mercolli, L., Pyka, T., Prenosil, G., Shi, K., Rominger, A., & Afshar-Oromieh, A. (2023, March 09). Large language models (LLM) and ChatGPT: what will the impact on nuclear medicine be? *European Journal of Nuclear Medicine and Molecular Imaging, 50, https://doi.org/10.1007/s00259-023-06172-w*, pp. 1549-1552.

Arcas, B. A. (2022, May 1). Do large language models understand us? *Daedalus, 151(2), doi: https://doi.org/10.1162/daed_a_01909*, pp. 183-197.

Asadi, A. R. (2023, December 26). LLMs in Design Thinking: Autoethnographic Insights and Design Implications. *Proceedings of the 2023 5th World Symposium on Software Engineering. https://doi.org/10.1145/3631991.3631999* (pp. 55-60). Tokyo: Association for Computing Machinery.

Azungah, T. (2018, November 30). Qualitative research: deductive and inductive approaches to data analysis. *Qualitative Research Journal, 18, 4, https://doi.org/10.1108/QRJ-D-18-00035*, pp. 383-400.

Basit, T. (2010, December 02). Manual or electronic? The role. *Educational Research, 45, 2, https://doi-org.ezproxy2.utwente.nl/10.1080/0013188032000133548*, pp. 143-154.

Bedekar, M. A., Pareek, M., Choudhuri, S. S., Abhishek, P. F., Jhurani, J., & Shah, R. (2024). AI in Mergers and Acquisitions: Analyzing the Effectiveness of Artificial Intelligence in Due Diligence. *2024 International Conference on Knowledge Engineering and Communication Systems, DOI: 10.1109/ICKECS61492.2024.10616599* (pp. 1-5). Chikkaballapur, India: IEEE.

Bellomo, S., Zhang, S., Ivers, J., Cohen, J., & Ozkaya, I. (2023, November, November). *Assessing Opportunities for LLMs in Software Engineering and Acquisition.* Pittsburgh: Carnegie Mellon University, Software Engineering Institute, DOI: 10.58012/m3hj-6w28.

Benson, P. (2012, November 05). Qualitative methods: Overview. *The Encyclopedia Of Applied Linguistics, https://doi.org/10.1002/9781405198431.wbeal0977*.

Borger, J. G., Ng, A. P., Anderton, H., Ashdown, G. W., Auld, M., Blewitt, M. E., . . . Naik, S. H. (2023, September 18). Artificial intelligence takes center stage: exploring the capabilities and implications of ChatGPT and other AI-assisted technologies in scientific research and education. *Immunology And Cell Biology, 101(10)*, pp. 923-935.

Brath, R., Bradley, A., & Jonker, D. (2024, September 10). Strategic management analysis: from data to strategy diagram by LLM. *arXiv:2409.06643, https://doi.org/10.48550/arXiv.2409.06643*.

Braun, V., & Clarke, V. (2006, January 1). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3(2), DOI:10.1191/1478088706qp063oa*, pp. 77-101.

Bristowe, K., Selman, L., & Murtagh, F. (2015, September 01). Qualitative research methods in renal medicine: an introduction. *Nephrology, dialysis, transplantation: official* *publication of the European Dialysis and Transplant Association - European Renal Association, 30(9), https://doi.org/10.1093/ndt/gfu410*, pp. 1424-1431.

Cao, L. (2022, February 03). AI in Finance: Challenges, Techniques, and Opportunities. *ACM Computing Surveys, 55(3), https://doi.org/10.1145/3502289*, pp. 1-38.

Chen, C., & Shu, K. (2023, September 25). Can LLM-Generated Misinformation Be Detected?. *arXiv:2309.13788, https://doi.org/10.48550/arXiv.2309.13788.*

Chen, C., & Shu, K. (2024, August 1). Combating misinformation in the age of LLMs: Opportunities and challenges. *AI Magazine, https://doi.org/10.1002/aaai.12188*.

Chen, S., Zhao, Y., Zhang, J., Chern, I., Gao, S., Liu, P., & He, J. (2023, Oktober 1). FELM: Benchmarking Factuality Evaluation of Large Language Models. *arXiv:2310.00741, https://doi.org/10.48550/arXiv.2310.00741*.

Cockburn, I. M., Henderson, R., & Stern, S. (2019). The impact of artificial intelligence on innovation: An Exploratory Analysis. In J. G. Ajay Agrawal, *The economics of Artificial Intelligence* (pp. 115-146). Chicago: University of Chicago Press.

Côté, J., Salmela, J., Baria, A., & Russell, S. (1993, June). Organizing and Interpreting Unstructured Qualitative Data. *IEEE Transactions on Signal Processing, 7, 2, https://doi.org/10.1123/TSP.7.2.127*, pp. 127-137.

Courteau, L. (2020, June 1). The information content of Fairness Opinions in M&A: Evidence from Italy. *Financial reporting, 1, https://doi.org/10.3280/fr2020-001001*, pp. 5-46.

Craig, L., Laskowski, N., & Tucci, L. (2024, October). *What is AI? Artificial Intelligence explained*. Retrieved from https://www.techtarget.com/searchenterpriseai/definition/AI-Artificial-Intelligence

Darayseh, M., & Alsharari, N. M. (2023, July 10). Determinants of merger and acquisition in the banking sector: an empirical study. *Meditari Accountancy Research, 31(4), DOI:10.1108/MEDAR-09-2021-1449*, pp. 1093-1108.

Davis, K. (1995). Qualitative Theory and Methods in Applied Linguistics Research. *TESOL Quarterly, 29(3), Autumn*, pp. 427-453.

del Río Carballo, P., González, F. J., & Sumaza, C. R. (2023, July). An Analysis of Mergers between Entities in the Sub-Sector of Intellectual and Developmental Disabilities in Spain. *Siglo Cero, https://doi.org/10.14201/scero.29546*, pp. 73-92.

Dillon, D. R. (2012, November 05). Grounded Theory and Qualitative Research. *The encyclopedia of applied linguistics,* , pp. 1-10.

Douglas, M. (2023, July 11). Large Language Models. *arXiv:2307.05782, https://doi.org/10.48550/arXiv.2307.05782*.

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., . . . Kronemann, B. (2021, April). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management, 57(101994), https://doi.org/10.1016/j.ijinfomgt.2019.08.002*.

Ergen, M. (2019). What is Artificial Intelligence? Technical Considerations and Future Perception. *The Anatolian Journal of Cardiology, 22(2), DOI:10.14744/AnatolJCardiol.2019.79091*, pp. 5-7.

Gao, T., Yen, H., Yu, J., & Chen, D. (2023, May 24). Enabling Large Language Models to Generate Text with Citations. *arXiv:2305.14627, https://doi.org/10.48550/arXiv.2305.14627.*

Gherghina, S. C. (2021, January 21). Corporate Finance. *Journal of Risk and Financial Management, 14(2), https://doi.org/10.3390/jrfm14020044*.

Gomes, M. d. (2018). *Role of Investment Banks in the M&A Process: Internship at BNP Paribas.* Carcavelos: NOVA – School of Business and Economics.

Harding, J., D'Alessandro, W., Laskowski, N. G., & Long, R. (2023, July 21). AI language models cannot replace human research participants. *AI & Society, 39*, pp. 2603-2605.

He Wang, Y. Z. (2023). Combination of Artificial Intelligence with Mergers and Acquisitions. *International Conference on Innovations in Marketing and Investment Management* (pp. 235-241). Tianjin City, China: BCP Business & Management.

Head, C. B., Jasper, P., McConnachie, M., Raftree, L., & Higdon, G. (2023, November 03). Large language model applications for evaluation: Opportunities and ethical implications. *New Directions for Evaluation, 2023, 178-179, https://doi.org/10.1002/ev.20556*, pp. 33-46.

Hennigan, K. M., Flay, B. R., & Haag, R. A. (1979). Clarification of Concepts and Terms Commonly used in Evaluative Research. In R. E. Klein, M. S. Read, H. W. Riecken, J. A. Brown, Jr., A. Pradilla, & C. H. Daza, *Evaluating the Impact of Nutrition and Health Programs, https://doi.org/10.1007/978-1-4684-3489-7_13* (pp. 387-432). Boston: Springer.

Hennigen, L. T., Shen, S., Nrusimha, A., Gapp, B., Sontag, D., & Kim, Y. (2024, April 15). Towards Verifiable Text Generation with Symbolic References. *arXiv:2311.09188, https://doi.org/10.48550/arXiv.2311.09188*.

Hsieh, H., & Shannon, S. E. (2005, November). Three Approaches to Qualitative Content Analysis. *Qualitative Health Research, 15(9)*, pp. 1277-1288.

Kalogirou, S. A. (2003, July 11). Artificial intelligence for the Modeling and Control of Combustion Processes: A review. *Progress in Energy and Combustion Science, 29(6), https://doi.org/10.1016/S0360-1285(03)00058-3*, pp. 515-566.

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., . . . Seidel, T. (2023, April). ChatGPT for good? On opportunities and challenges of large language. *Learning and Individual Differences, 103(102274), https://doi.org/10.1016/j.lindif.2023.102274*.

Ketokivi, M., & Choi, T. (2014, July). Renaissance of case research as a scientific method. *Journal of Operations Management, 32, 5, https://doi.org/10.1016/j.jom.2014.03.004*, pp. 232-240.

Kuckartz, U. (2014). Three basic methods of Qualitative text analysis. In U. Kuckartz, *Qualitative Text Analysis: A Guide to Methods, Practice & Using Software* (pp. 65-120). SAGE.

Kuckartz, U. (2019, April 27). Qualitative Text analysis: a Systematic approach. *ICME-13 monographs*, pp. 181-197.

Lesterhuis, M., Bouwer, R., Daal, T., Donche, V., & Maeyer, S. (2022, May 13). Validity of Comparative Judgment Scores: How Assessors Evaluate Aspects of Text Quality When Comparing Argumentative Texts. *Sec. Assessment, Testing and Applied Measurement, 7, https://doi.org/10.3389/feduc.2022.823895*.

Li, Y. (2023). Application of AI Technology in Mergers & Acquisitions and Its Regulation. *Advances in Economics, Management and Political Sciences, 11, (1), https://doi.org/10.54254/2754-1169/11/20230502*, (pp. 32-38). Guangzhou, China.

Li, Y. Y., Wang, S. S., Ding, H. H., & Chen, H. H. (2023, November 25). Large Language Models in Finance: A Survey. *Proceedings of the Fourth ACM International Conference on AI in Finance, ISBN: 979-8-4007-0240-2.* New York: Association for Computing Machinery.

Lin, Y., Kumar, A., Zhang, W., Chang, N., Zakir, M., Apte, R., . . . Jang, J. (2024, January 30). Applications of Large Language Models in Data Processing: Innovative Approaches to Segmenting and Renewing Information. *arXiv:2311.16267, https://doi.org/10.48550/arXiv.2311.16267*.

Liu, X. Y., Zhang, H., Xu, Y., Lei, X., Lai, H., Gu, Y., . . . Tang, J. (2023, August 7). AgentBench: Evaluating LLMs as Agents. *arXiv:2308.03688, https://doi.org/10.48550/arXiv.2308.03688.*

Magnani, G., & Gioia, D. (2023, April). Using the Gioia Methodology in international business and entrepreneurship research. *International Business Review, 32, 2, https://doi.org/10.1016/j.ibusrev.2022.102097*.

Mang'ana, K. m., Hokororo, S. J., & Ndyetabula, D. W. (2024, January 1). An Investigation of the Extent of Implementation of the Financial Management Practices of Agri-SMEs in developing countries: Evidence from Tanzania. *Sustainable Technology and Entrepreneurship, 3(1), https://doi.org/10.1016/j.stae.2023.100049*.

Martin, R. L. (2016, June 1). M&A: The One Thing You Need to Get Right. *Harvard Business Review, 94(6), 12. https://dialnet.unirioja.es/servlet/articulo?codigo=5486047*.

Meyer, B. J. (2003, July). Text Coherence and Readability. *Topics in Language Disorders, 23, 3*, pp. 204-224.

Nejjar, M., Zacharias, L., Stiehle, F., & Weber, I. (2023, November 28). LLMs for Science: Usage for Code Generation and Data Analysis. *arXiv:2311.16733, https://doi.org/10.48550/arXiv.2311.16733.*

Nogueira, N. V., & de Castro, L. R. (2019, December 6). Effects of ownership structure on the mergers and acquisitions decisions in Brazilian firms. *RAUSP Management Journal, 55, 2, https://doi.org/10.1108/RAUSP-11-2018-0124*, pp. 227-245.

Pallathadka, H., Ramirez-Asís, E., Loli-Poma, T. P., Kaliyaperumal, K., Ventayen, R. J., & Naved, M. (2023). Applications of artificial intelligence in business management, e-commerce and finance. *Materials Today: Proceedings, 80(3), https://doi.org/10.1016/j.matpr.2021.06.419*, pp. 2610-2613.

Patton, M. Q. (2015). *Qualitative research & evaluation methods.* Saint Paul: SAGE Publications, Inc.

Rasnayaka, S., Wang, G., Shariffdeen, R., & Iyer, G. N. (2024, January 29). An empirical study on usage and perceptions of llms in a software engineering project. *arXiv:2401.16186, https://doi.org/10.1145/3643795.3648379*, pp. 111-118.

Rien, A. (2018). *The Art of AI - The Impact of Artificial Intelligence on the Merger and Acquisition Strategy.* Lisbon: Master's thesis, Universidade Catolica Portuguesa.

Roberts, C. W. (2000, August 01). A Conceptual Framework for Quantitative Text Analysis. *Quality and Quantity, 34*, pp. 259-274.

Ryu, C., Lee, S., Pang, S., Choi, C., Choi, H., Min, M., & Sohn, J.-Y. (2023, December). Retrieval-based Evaluation for LLMs: A Case Study in Korean Legal QA. *Proceedings of the Natural Legal Language Processing Workshop, https://doi.org/10.18653/v1/2023.nllp-1.13* (pp. 132-137). Singapore: Association for Computational Linguistics.

Sadler, D. (1989, June). Formative assessment and the design of instructional systems. *Instructional Science, 18, https://doi.org/10.1007/BF00117714*, pp. 119-144.

Saha, T., Ganguly, D., Saha, S., & Mitra, P. (2023). Workshop On Large Language Models' Interpretability and Trustworthiness (LLMIT). *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management, https://doi.org/10.1145/3583780.3615311* (pp. 5290-5293). Birmingham, United Kingdom: Association for Computing Machinery.

Schilling, J. (2006). On the Pragmatics of Qualitative Assessment Designing the Process for Content Analysis. *European Journal of Psychological Assessment, 22(1), https://psycnet.apa.org/doi/10.1027/1015-5759.22.1.28*, pp. 28-37.

Seers, K. (2011, December 19). Qualitative data analysis. *Evidence Based Nursing, 15(1), https://doi.org/10.1136/ebnurs.2011.100352*, pp. 2-2.

Shah, P., Mayer, R. E., & Hegarty, M. (1999). Graphs as aids to knowledge construction: Signaling techniques for guiding the process of graph comprehension. *Journal of Educational Psychology, 91, https://doi.org/10.1037/0022-0663.91.4.690*, pp. 690-702.

Shao, J., Lou, Z., Wang, C., Mao, J., & Ye, A. (2021, October 1). The impact of artificial intelligence (AI) finance on financing constraints of non-SOE firms in emerging markets. *International Journal of Emerging Markets, 17(4), https://doi.org/10.1108/IJOEM-02-2021-0299*, pp. 930-944.

Shi, Z., & Zheng, N. (2006, September). Progress and challenge of artificial intelligence. *Journal of Computer Science and Technology*, pp. 810-822.

Sorin, M., Glicksberg, P., Barash, M., Konen, M., Nadkarni, M., & Klang, M. (2023, November 4). Applications of Large Language Models (LLMs) in Breast Cancer Care. *medRxiv (2023), https://doi.org/10.1101/2023.11.04.23298081*.

Sun, Z. (2023, March 16). A Short Survey of Viewing Large Language Models in Legal Aspect. *arXiv:2303.09136, https://doi.org/10.48550/arXiv.2303.09136*.

Tikkakoski, E. (2018). *Emotions related to mergers and acquisitions.* Turku: Turku School of Economics.

Topsakal, O., & Akıncı, T. Ç. (2023, July). Creating large language model applications Utilizing LangChain: A primer on developing LLM apps fast. *International Conference on Applied Engineering and Natural Sciences, 1(1), DOI: 10.59287/icaens.1127*, pp. 1050-1056.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., . . . Polosukhin, I. (2017, June 12). Attention is All You Need. *arXiv:1706.03762, https://arxiv.org/pdf/1706.03762v5*, pp. 5998–6008.

Veen, D., Uden, C., Blankemeier, L., Delbrouck, J., Aali, A., Blüthgen, C., . . . Chaudhari, A. (2023, September 14). Clinical Text Summarization: Adapting Large Language Models Can Outperform Human Experts. *Research Square, https://doi.org/10.21203/rs.3.rs-3483777/v1*.

Wang, H., & Zhou, Y. (2023, February 2). Combination of Artificial Intelligence with Mergers and Acquisitions. *BCP Business & Management, 39, https://doi.org/10.54691/bcpbm.v39i.4069*, pp. 235-241.

West, R., Marsden, J., Humphreys, K., & Darke, S. (2018, June 19). A revised checklist for writing up research reports. *Addiction, 113(9), https://doi.org/10.1111/add.14269*, pp. 1567–1570.

White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., . . . Schmidt, D. C. (2023, February 21). A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT. *arXiv:2302.11382, https://doi.org/10.48550/arXiv.2302.11382*.

Witteveen, S., & Andrews, M. (2019, November 1). Paraphrasing with Large Language Models. *arXiv:1911.09661*, pp. 215-220.

Yamauchi, R., Sonoda, S., Sannai, A., & Kumagai, W. (2023, September 21). LPML: LLM-Prompting Markup Language for Mathematical Reasoning. *arXiv:2309.13078, https://doi.org/10.48550/arXiv.2309.13078*.

Yang, J., Jin, H., Tang, R., Han, X., Feng, Q., Jiang, H., . . . Hu, X. (2024, April 26). Harnessing the Power of LLMs in Practice: A Survey on ChatGPT and Beyond. *ACM Transactions On Knowledge Discovery From Data, 18(6), https://doi.org/10.1145/3649506*, pp. 1-32.

Yao, Y., Duan, J., Xu, K., Cai, Y., Sun, E., & Zhang. (2024, March 20). A Survey on Large Language Model (LLM) Security and Privacy: The Good, the Bad, and the Ugly. *arXiv:2312.02003, https://doi.org/10.48550/arXiv.2312.02003*.

Ye, W., Ou, M., Li, T., Chen, Y., Ma, X., YangGong, Y., . . . Zhao, J. (2023, August 30). Assessing Hidden Risks of LLMs: An Empirical Study on Robustness, Consistency, and Credibility. *arXiv:2305.10235, https://doi.org/10.48550/arXiv.2305.10235*.

Zhao, H., Liu, Z., Wu, Z., Li, Y., Yang, T., Shu, P., . . . Liu, T. (2024, January 22). Revolutionizing finance with llms: An overview of applications and insights. *arXiv:2401.11641, https://doi.org/10.48550/arXiv.2401.11641*.

Appendix

Appendix A – Assessment Format

Qualitative

1. What is your impression of this analysis?
2. How relevant is the text that is written in this analysis?
3. What is/are the strongest parts in this text?
4. What is/are the weakest parts in this text?
5. How complete is the text?
6. What are important factors that need to be accounted for, when writing a text?
7. Is the text factual correct, compared to the original text that was used in the actual IM?

Quantitative

8. Can you assess the clarity of this text on a scale from 1 to 10?

← Bad					Good →				

9. Can you assess the relevance of this text on a scale from 1 to 10?

← Bad					Good →				

10. Can you assess the structure of this text on a scale from 1 to 10?

← Bad					Good →				

11. What is your overall judgment of the text on a scale from 1 to 10?

← Bad					Good →				

Estimated time

1. How much time did you spend assessing this text?

2. How much time would you expect to spend to make this text applicable for an IM?

Appendix B – Interview Guide M&A Consultants

1. What working experience do you have?
    a. What companies have you worked for in the past?
    b. What is your relevant working experience, and in what roles?
    c. How long have you been working at Taurus, and in what roles?
    d. What is your current position at Taurus?
2. What kind of M&A transactions do you usually work on?
    a. What is the size of the companies that you guide, in terms of EBITDA and staff?
    b. What is the range of the transaction size for the companies that you guide?
    c. What kind of business sectors do you usually guide companies in, regarding the M&A process?
3. How do you do your research when it comes to the internal analysis of a company?
    a. What sources do you use for your research?
    b. How much time does this research take?
4. How do you do your research when it comes to the external analysis of a company?
    a. What sources do you use for your research?
    b. How much time does this research take?
5. How much time does the writing process take, when it comes to the internal analysis of a company?
6. How much time does the writing process take, when it comes to the external analysis of a company?
7. What is your experience with LLMs?
    a. Have you used any LLMs?
    b. What LLMs do you use?
    c. How helpful do you think LLMs are?
8. How can LLMs assist an M&A consultant with their work?
    a. How can it assist in the internal and external analysis of a company?
    b. Are there tasks for an M&A consultant that are repetitive, that could be replaced or assisted by LLMs to improve efficiency?

c. What are those tasks?

9. Which components, with regard to the internal and external analysis in the M&A process, do you expect to run more efficiently, by using LLMs?
10. Which components, with regard to the internal and external analysis in the M&A process, cannot be replaced by LLMs?

Appendix C – Interview Guide IT / AI Expert

1. What working experience do you have?

    e. What companies have you worked for in the past?

    f. What is your relevant working experience, and in what roles?

    g. How long have you been in IT, and in what roles?

    h. What is your current position?

2. What is your experience with LLMs?

    a. Have you used any LLMs?

    b. What LLMs do you use?

    c. How helpful do you think LLMs are?

3. What are the most important capabilities of LLMs that could be useful for an M&A consultant?

    d. How can it assist in the internal and external analysis of a company?

4. Which subsections, with regard to the internal and external analysis in the M&A process, do you expect to run more efficiently, by using LLMs?

5. Which subsections, with regard to the internal and external analysis in the M&A process, cannot be replaced by LLMs?

6. What is the best way to prompt an LLM to let it generate a useful text for you?

    a. What are things that you should account for, when writing a prompt?

7. What are the most important dangers of using LLMs in corporate finance?

8. What are your expectations for the future of LLMs and its applications in Corporate finance?

Appendix D - Coding scheme

Gioia method for coding interviews

1st-Order Concepts	2nd-Order themes	Aggregate dimensions
- I request a lot of information from the client to perform an internal analysis. - I often conduct an interview with the client to understand how their administrative organization is structured, the composition of their workforce, and how they define their strategy, both internally and externally. - You walk through the company with the client on all business fronts, relying heavily on qualitative information. - We use various databases, which contains a lot of information about competitors of the company in question and research reports on the market the company operates in. - We try our best to identify the industry trends and developments. - Market development involves scraping a lot of data, both nationally and internationally, across Europe and worldwide, sometimes even regionally. That's something AI can handle more easily. - For every project, you need to do a market analysis. - we view an internal analysis more as an analysis within an external environment. ... It's more about comparing the company to its external environment. - It's more understanding why this company is outperforming its competitors or other companies in the market. - you need to have enough confidence that you did the research yourself. I compare it a bit to using a calculator: you can only use a calculator effectively when you can roughly estimate the size of the result yourself. So if you have no idea that 100x100 should come out somewhere around 10,000, and you enter 99x102 and get 102,000 as a result, you won't be able to correct it. - The internal analysis is partly qualitative, where we discuss the organization, the equity story, where the company stands, its legacy, and milestones. - Requesting information and also through Q&A sessions with the entrepreneur.	- Emphasis on personal expertise and validation in research - Integrating internal analysis with external environmental context - Extensive external data sources - Qualitative exploration of organizational identity	Relevance

Gioia method for coding interviews

1st-Order Concepts

- The most important thing is to have extensive conversations with the client to accurately identify various aspects of their business.
- Internal analysis is necessary to validate the external findings and make a meaningful comparison.
- A conversation with the management helps you identify where you need to ask more questions, gather additional information, and document insights that may not be relevant to a market competitor but are crucial for an investor.
- AI can identify patterns and work predictively. Someone would operate the machine manually, but now a robot does it. The operator still needs to set it up and conduct quality checks. I think that it works the same for AI.
- LLMs would struggle with internal analysis such as understanding how a company is organized, which employees work in what ways and identifying key players, if it's based purely on data sources, sometimes public and sometimes less so, it's unlikely to produce good answers.
- The information is often very refined and not typically available in public sources.
- Research is necessary, but some tasks can be handled by an LLM.
- A language model is just a very good predictor and doesn't always understand the context.
- So, there is still a challenge in making sure you verify it. Also, in rewriting or ensuring that the context and the emphasis you place are correct. Yes, you definitely need to do that.
- I believe that you can have 90% of your work done very well, but you'll need to add or correct 10 to 20% yourself to make the difference.
- In two years, a lot can change in an external analysis. So yes, you can get quite far, but up-to-date information, right up to today, won't be completely accurate.
- If you ask for a sentence from one A4 page and ask it to analyze and summarize that, it does that well. But if you throw 3,000 A4 pages at it, how do you know it's considering all 3,000? How have you validated that in your prompt? You can scan that one A4 page as a human to see what's in it. But we can't scan 3,000 A4 pages. So, if you throw 3,000 A4 pages into the LLM in your prompt, and the prompt thinks, "This is going well, I'll have something for you in a couple of minutes," how do you know the prompt didn't stop at page 100?
- AI just provides a lot of data, and then you have to validate it.

2nd-Order themes

- Client interaction for in-depth understanding of information
- Necessity of internal analysis for validating external data
- Limitations of LLMs in capturing contextual and confidential Information
- Challenges in handling and validation of LLM-generated texts
- Verification and contextualization in LLM-generated texts

Aggregate dimensions

Completeness

Gioia method for coding interviews		
1st-Order Concepts	**2nd-Order themes**	**Aggregate dimensions**
- You always need to validate whether the output of these interviews can be made objective. - You need to challenge those expectations to ensure they are realistic and the analysis holds up. - Objectivity can be achieved by comparing these expectations against market information and historical performance data to see if the anticipated growth is sustainable. - Research on the company itself remains essential - Large corporations usually have better reporting, even internally, but SMEs often don't. The information is often still within the entrepreneur or the family, so you need them to even get the story on paper. - You still need to verify the LLM output because it's risky to blindly trust a suggestion. - Internal analysis could be similar, though with companies like ours, where information often resides within the company, it's still a bit tricky to extract that. - if you really want to implement it in your company, then you need such company-specific knowledge. That's where you encounter a different aspect, which a standard LLM doesn't handle so well. Because, in the end, it's just a prediction mechanism, so it includes a lot of data. But most of it is standard data, and what makes the difference, that's not always as refined in the current LLMs. - If there's an AI engine like GPT-4 behind it, then the data is up to 2022. That's not very helpful now that we're in 2024. So, you have to realize that it's important to know what your goal is and what you want to do with your data, especially also considering what your sources are. - Because you can produce results so quickly with a good language model, critical thinking might be forgotten. - you could say to the LLM, "As you've written it in the summary, that's great, but could you also tell me at the end of your conclusion how many words you used to analyze?" Now, if you have this document in Word with 3,000 pages, you can see that it contains 10,000 or 20,000 words. You can see that in the document's properties. So, if it says, "I used 2,000 words," you might think that's odd because that's too few. So, you also need to consider the integrity of your analysis. - I think one of the main dangers is that the user might become complacent, and you may not have a clear way to verify what is actually true. - If something happens in an analysis and you ask an LLM to work its magic with its data, we assume it's doing it correctly. This increases dependence. If you're 90% correct, you probably won't doubt the remaining 10%, so you won't question it. You become too lazy because the speed is so great that you also need to think about how to validate that high speed with the data you're working with.	- Validation and objectivity in analysis - Limitations of LLMs in handling company-specific and up-to-date information - Challenges in extracting and verifying internal company information - Understanding key information of companies - Human oversight to ensure integrity and accuracy of LLM-generated texts	Factual correctness

Gioia method for coding interviews

1st-Order Concepts	2nd-Order themes	Aggregate dimensions
- LLMs are still mainly a summarizing tool, where you only know if the summary makes sense if you already know the subject well or partially. - in the short term, it can at most accelerate writing or summarizing tasks. - I sometimes put SPA's in it to see what it extracts, which works well, but you have to be good at controlling the output. It's still somewhat generic in what it can do. The human touch is still important to control the output. - It can interpret and summarize well, but you still need to control the output. - LLMs can help M&A consultants with quick data analysis, fast summaries of large text chunks, and very directional insights, think about this, consider that. - You have models that are better with language, in Dutch, in English. So, you can use those better for interpretation. Gemini has made a huge leap in that. - So, especially when it comes to marketing texts and rewriting existing texts, yes, it's really perfect for that, and I think it's also perfect for completing texts. You can use it very well for that. So, I think if you're uninspired, thinking, "How in the world am I going to do this?" Yes, you can use a standard LLM very well for that. They are excellent text generators. - If you want to create an analysis for a company or make a nice introduction text for a company, or you want a backup procedure, it has all kinds of existing prompt texts that you can extract. So, in the open structure, you can get quite a lot out of it. - For the Dutch language, the functionality with Gemini would currently yield better results. Gemini is more advanced in the Dutch language than ChatGPT. - LLMs are a language magician that doesn't understand language or context	- LLMs are an effective summarizing and interpretation tool - LLMs are language magician - LLMs can be used to rewrite texts in a more readable fashion - Utility of LLMs in providing templates and texts for business documentation	Text clarity/ structure

Gioia method for coding interviews

1st-Order Concepts

- I think you spend about 30 hours on research to gather the necessary information. Then you spend another 30 hours on writing for both parts, so about 60 hours total for both research and writing, for both the internal and external analyses.
- The workload for a transaction can sometimes be more than four times the workload of another transaction, so the number of transactions may not fully reflect the effort involved.
- If you see a text that is generated by an LLM, and scan the sources used, double-checking them via Google or other sources, I think it would take roughly two hours to verify.
- in the short term, it can at most accelerate writing or summarizing tasks.
- A lot of the analysis and further analysis only happens after you've completed part of the analysis and written part of it. So, purely writing an IM can be done in a day. I mean, there aren't that many words on paper. But from the first draft to the final version, it can take about a month.
- In general, if you consider 30 hours for internal analysis, 40 hours for external analysis, and 40 to 50 hours for financial analysis, that's about 100 to 120 hours. You're likely to spend more time, if you're also including writing.
- You'll have more time for the good things, spending more on customer experience, service, and communication. You can serve your clients better.
- You could save between 40 and 70 percent of your time over the year. You could then spend that time on design, storyline, etc.
- Just checking and screening to make sure there's nothing strange in it, that's 10% of the work, compared to 90% of your work just to get a text on paper in the first place.
- a lot of the work involves analyzing the files you receive. The great thing is that AI can analyze this much faster.
- So, AI will have an impact on automation, on marketing, on sales. I think you'll need half as many people, and you'll need very different kinds of people. But we have a labor shortage, so there will still be a need if you want to grow. But yes, I think the impact will be enormous.
- Productivity becomes much shorter, much faster. Therefore, the purpose of the work also changes.
- If you really want to implement it in your company, then you need such company-specific knowledge.

2nd-Order themes

- Time spent for analyzing
- Time spent for writing
- Saving time with the use of LLMs
- LLMs will help automate tasks

Aggregate dimensions

Time spent

Appendix E – Interview A- Taurus expert.

What is your work experience?

I started at ING Bank in 2005 and worked there until 2020, so I spent 15 years at ING in various roles. I was mostly involved in corporate services, serving business clients. For about 7 or 8 years, I focused primarily on real estate financing. Additionally, I handled a lot of acquisition financing within ING Bank. Four years ago, I made the switch to Taurus, and now I am one of the partners here at Taurus, working with a team that will soon have 25 colleagues. At Taurus, we handle around 30-40 sales transactions per year. This involves everything from the initial strategic discussions with the client about the company's growth and next steps to the final stage, 9-12 months later, when the shares are sold and transferred at the notary.

Did you also work on M&A transactions at ING?

Yes, but primarily on the financing side. The financing side is actually the final stage in the sales process. It follows everything that happens beforehand, such as talks with potential buyers, negotiations, determining the sales structure, and then, in the final stage, the transaction needs to be financed. This can be partly through the company's own capital, but it often involves bank financing as well. I was involved in that final stage from ING's side.

When you joined Taurus, what were your roles there?

At Taurus, my roles have mainly focused on commercial activities, especially trying to secure new transactions for Taurus. This includes visiting clients, conducting many acquisition talks, and attending numerous networking events to attract business. Engaging with clients who are in the process of selling and guiding them through that process is still a significant part of my job. Once a client has signed a sales mandate, I also take on the role of project leader for the transaction. Initially, I serve as the main point of contact for the client and guide the team in organizing all the information, eventually compiling it into an information memorandum. I also lead discussions and negotiations with potential buyers.

What kind of M&A transactions do you typically advise on?

Generally, I advise on sales transactions. I also have a few clients with a buy-and-build strategy, meaning they aim to grow through acquisitions and have given me the mandate to look for potential companies that fit well with their business. But this is the case for only about two clients. Beyond that, I handle around 10-15 sales transactions a year. In my case, these often involve corporate services, such as staffing agencies, IT companies, e-commerce platforms, and similar businesses.

How large are the companies you advise and guide, in terms of EBITDA and staff?

This varies greatly. For example, I currently have a client with 500 employees, €120 million in revenue, and €30 million in EBITDA. But I also advise a company with €5 million in revenue, 5 employees, and €500,000 in EBITDA. It is a broad spectrum, but most clients fall in the €2-10 million EBITDA range.

What is the size of the transactions for the companies you have advised?

These range from transactions worth over €200 million to others with a valuation of around €5 million.

How do you conduct research when performing an internal analysis of a company?

I request a lot of information from the client, including annual reports and audit files, to perform an internal analysis. Additionally, I often conduct an interview with the client to understand how their administrative organization is structured, the composition of their workforce, and how they define their strategy, both internally and externally. Essentially, you walk through the company with the client on all business fronts, relying heavily on qualitative information.

Are there any other sources you use for this?

No, for an internal analysis, not really.

How do you conduct research for an external analysis of a company?

We use various databases, including Gain.Pro, which contains a lot of information about competitors of the company in question and research reports on the market the company operates in. Additionally, there is a lot of Googling involved to identify the industry trends and developments. So, much of it is done through simple Googling and using the right search terms.

How much time does the writing process take when it comes to the internal analysis of a company?

I think you spend about 30 hours on research to gather the necessary information. Then you spend another 30 hours on writing for both parts, so about 60 hours total for both research and writing, for both the internal and external analyses.

What is your experience with large language models?

Hardly any. I think I've used ChatGPT, which is indeed a large language model, four or five times. So far, I've left the use of ChatGPT to others.

How useful do you find LLMs?

From what I've seen, mostly from colleagues, LLMs can indeed be very useful. Especially for an external analysis, if you use the right terminology, you see a whole lot of information coming in. So, it is very useful, but I can imagine, and this is also what I've read about LLMs, that you really need to be cautious, ensuring that what comes out is not rubbish. So, checking whether it is misinformation is important.

How could LLMs assist an M&A consultant in their work?

I think that, especially for external analysis, you can quickly gather and organize a lot of information. So, it can save a lot of time in that area. It also forces you to think more deeply about the information you have gathered and to take a step further to better understand the market.

Are there tasks for an M&A consultant that are repetitive and could be replaced or assisted by LLMs to improve efficiency?

Yes, repetitive tasks include market analysis. For every project, you need to do a market analysis. For example, in the staffing industry, there is a generic market analysis covering aspects like monthly revenue trends and hourly developments. Staffing agencies can often be categorized into segments, like those focusing on blue-collar workers, really LBO, MBO level, where you can zoom in specifically. On the other hand, there are also white-collar staffing agencies where this applies.

Which sub-sections in the information memorandum related to internal and external analysis in the M&A process do you expect could be made more efficient by using large language models?

Well, as I said, the sub-sections would be the external analysis, including market analysis and competitor analysis, and possibly a customer analysis. You could organize a client analysis in that regard. For internal analysis, you could possibly arrange a benchmark analysis by identifying some competitors and comparing them against the company you are selling. But also, potentially for an internal analysis of how the supply chain is organized for this company and where the company is positioned in the supply chain, and what opportunities and possibilities exist within the supply chain. You could also ask an LLM for potential candidates or buyers, but the challenge with LLMs is that they cannot really see inside the company. So, organizing a good internal analysis might be difficult. For example, if you ask what the optimal structure for an administrative organization is for a technical staffing agency, I have no idea what the answer would be. But you might be able to ask an LLM about an optimal situation, which you can then test against the company. For example, in the internal analysis, you also have a section on housing and personnel, which is harder to do with an LLM.

Which sub-sections related to internal and external analysis could not be replaced by an LLM?

When it comes to the administrative organization and internal organization, I think it is very difficult to do that via an LLM because it is really company-specific and that information is not publicly available. So, in my view, you could only test it against a benchmark in the segment rather than using it for that specific information. Any information about a company that has not been made publicly available cannot be run through an LLM, in my opinion. At most, you could do it via a private LLM for the company. We also have a lot of information stored in a data room. Perhaps in the future, it will be possible to apply an LLM to your data room and ask questions like, "How is the internal organization structured?" and then the LLM would write about that. This could be done with the information available in the data room. Beyond that, you would still need to research the company itself, as otherwise, you would not be able to verify the LLM's output.

How much time do you think it would take to verify the accuracy of information generated by an LLM?

It is hard for me to estimate because I've never done it myself. But if you see a text and scan the sources used, double-checking them via Google or other sources, I think it would take roughly two hours to verify.

Appendix F – Interview B- Taurus expert.

What is your work experience? In 2016, after completing my MBA, I started working at Taurus, where I was the fourth employee at the time. So, I essentially began as a junior analyst. However, being in a team of four is quite different from being in a team of over twenty people, as we have now. I gradually moved up to a senior analyst role. Eventually, I became a project leader, back when we used that title instead of manager, which we now have. Finally, I became a partner early last year. So, I now have eight years of work experience since 2016, and I've gone through all the stages within Taurus.

What kind of education did you complete before this? I earned a bachelor's degree in international business with a minor in accountancy, and then completed a master's in business administration, majoring in financial management. My thesis focused on business valuation.

What types of M&A transactions do you typically advise on? Almost exclusively sell-side transactions. I tend to work more with younger entrepreneurs who are selling to private equity parties, rather than to competitors. So, I primarily handle pre-exit scenarios.

How many transactions do you generally handle annually? Between three and six. The workload for a transaction can sometimes be more than four times the workload of another transaction, so the number of transactions may not fully reflect the effort involved. I think the deal value is not always a perfect indicator either, but it falls somewhere in between. Typically, I handle between three and six transactions, with a total enterprise value ranging from 60 to 400 million euros annually.

How large are the companies you assist in terms of EBITDA and staff? There are significant differences. The smallest company in my current portfolio has an EBITDA of 1.5 million, while the largest has 6 million. However, last year I handled a company with an EBITDA of 30 million, so it varies widely.

Do you specialize in specific sectors? Not deliberately, but the market sometimes perceives you as a specialist in certain sectors. I've handled several transactions in e-commerce in the past, and in this small world, the market can start to view you as an e-commerce specialist. So, I do more e-commerce deals, although it is not something I consciously seek out, it just helps.

When it comes to the internal analysis of a company, how do you conduct your research? The most important thing is to have extensive conversations with the client to accurately identify various aspects of their business. It becomes easier as you gain more experience, gather more data points, and have more examples of how things work within other companies, allowing you to make better comparisons. With more experience in conducting internal analyses, you can more easily identify differences and similarities, and recognize what sets a company apart. I think "internal analysis" might be a somewhat tricky term. If we are strictly talking about the organization itself, our analysis might not go very far. I believe we view it more as an analysis within an external environment. So, it is more about understanding why this company is outperforming its competitors or other companies in the market. Why does this company achieve higher or lower margins than its competitors or peers in the market? It is more about comparing the company to its external environment.

What other sources do you use for your research? Besides interviews, which I believe are the most valuable as they provide clear directions, you always need to validate whether the output of these interviews can be made objective. For example, when a business owner expects 100% growth over the next three years and predicts a gross margin increase of 5, 6, or 7%, you need to challenge those expectations to ensure they are realistic and the analysis holds up. Objectivity can be achieved by comparing these expectations against market information and historical performance data to see if the anticipated growth is sustainable.

How much time does this research generally take? It can take anywhere from two weeks to two months. This largely depends on the situation. If a client shows a lot of ambition and clearly states that they want to grow their company three or four times in size over the next five years, there is much more evidence required to make those projections objective compared to a client who says they expect to continue doing what they have done over the past three years for the next three years as well. In that case, you mainly need to prove that there are no obstacles preventing them from doing the same in the coming years. This kind of analysis can be done in a few days or a week. However, when high expectations or promises are made, it requires a lot more substantiation, which could mean working with a strategy that helps determine the market's range for the next 5 to 10 years. So, it is not easy to give a definitive time frame.

When it comes to the external analysis of a company, how do you conduct your research? We might be mixing up internal and external analysis a bit here. What I’ve been talking about mostly relates to external analysis. You need to view the company within its environment of customers, suppliers, and competitors. Internal analysis is necessary to validate the external findings and make a meaningful comparison. If by internal analysis you mean the entrepreneur's expectations and their perspective on their company, then external knowledge is needed to understand whether those expectations and the current state of the company are realistic for the future. I do not see internal and external analysis as entirely separate processes; they should be considered together. An interview, which often involves speaking with management, is usually the first step to forming a market view from their perspective. This helps you identify where you need to ask more questions, gather additional information, and document insights that may not be relevant to a market competitor but are crucial for an investor. This ensures that everyone reading the documents starts with the same level of understanding. To achieve this, you first need to fully understand the company and the market yourself. The process always begins with a conversation with the entrepreneur, and from there, you determine what you need to support the analysis effectively.

The research on the company must always be done by yourself. Otherwise, you cannot actually verify whether AI has done it correctly. We are now using Gain.Pro as a resource to map out the competitive landscape and to see what kind of acquisitions have been made in the market, or in similar industries, that could be a good fit for this company. If you cite the AI as a kind of reference, saying, "I derived this from the Gain.Pro database," then it does not really matter whether you personally went through it all or whether AI did it for you. But you need to have enough confidence that you did the research yourself. I compare it a bit to using a calculator: you can only use a calculator effectively when you can roughly estimate the size of the result yourself. So if you have no idea that 100x100 should come out somewhere around 10,000, and you enter 99x102 and get 102,000 as a result, you will not be able to correct it. So, you can only use a calculator properly when you know what the result should roughly be. I see AI in a similar way. You can only use AI effectively when you have enough background knowledge to assess whether the conclusion it draws is close to the truth. If you use AI without that knowledge, you cannot validate whether what comes out is accurate or logical.

That is indeed why the research remains necessary. Research on the company itself remains essential, with AI being used only as a writing tool to help generate the text more easily. Is this purely from an IM perspective?

Yes, this is purely from the internal and external analysis of the IM. No, I agree with you on that. For now, I think that will certainly be the case in the coming year, although, yes, we also did not fully know how the internet would develop in the early stages. If you had to explain the internet to someone in the early 1990s, you might say, "Yes, you can send mail just as easily to your neighbor as to someone in Japan." Often, people would think, "Yeah, that is great, but I never send anything to Japan, and I can just drop it in the neighbor's mailbox, so I do not send that much mail." If the internet took over half of TNT Post's business, they'd be doing well. That perspective has shifted somewhat by now. Honestly, we do not know if AI might develop into something we cannot yet foresee. It would make sense that more is possible than we can currently imagine. But if I look purely at how AI is being used now, it is still mainly a summarizing tool, where you only know if the summary makes sense if you already know the subject well or partially. So, I am hesitant to predict its broader implications in the medium term, but I think that in the short term, it can at most accelerate writing or summarizing tasks. The focus is indeed on Large Language Models, like ChatGPT and Gemini, for example.

How much time does it take for the writing process per IM to put the internal and external analysis on paper? A lot of the analysis and further analysis only happens after you have completed part of the analysis and written part of it. So, purely writing an IM can be done in a day. I mean, there are not that many words on paper. But from the first draft to the final version, it can take about a month. This is mainly because you sit down with the client and discuss, "You distilled this from our conversation, you have supported some points well, but others not so much." You might come to the conclusion that, although you initially planned to support something in a certain way, the data does not actually back it up as hoped. So, it is really a collaboration with the client. This can take anywhere from a week to two months, depending on the situation. If your analysis is spot on right from the start, it can go much faster. If the company is straightforward with fewer complications, it also goes quicker. If the company is very complex, with perhaps five different revenue sources, then it is a different story, you might need to conduct five analyses instead of one.

Currently, I am also working with Employee 1, 2 and 3. We had a meeting with AI expert to discuss our goals. One of the goals is to improve the quality of the first draft of the IM that we send to the client, so that less interaction with the client is needed to get it to the final level. This means starting with a higher-quality IM from the beginning. Yes, that is a good initiative and a good goal. Honestly, you only know if it is effective once you have had that conversation with the client. If you have not fully understood what makes the company special in the initial discussions, then even with AI, you will end up with analyses that do not quite align with the client's perspective. So, a lot of it still depends on your own processing of whether you have truly captured the essence of the company during the conversation. Often, you realize after the first conversation that what you thought was a unique aspect of the company is actually slightly different after a second discussion with more examples. So, AI may not always be able to help if you cannot deliver the right input, so to speak.

What is your experience with Large Language Models (LLMs)? My experience is quite limited. It is more of an interest outside of work, where I experiment a bit, rather than something I use extensively in my professional life. So my experience is very limited. There are already numerous AI applications, and I believe AI is similar to the internet in that it offers various solutions to different problems. The internet is not just one site. The internet became significant because it offered many different solutions for many different problems, including many we did not even realize we needed before the internet existed. I think it will be the same with AI. Logically, I find applications like those in Excel relatively low-level compared to what is possible with AI. Applying AI sounds very interesting to us, much more so to perhaps use it for an external analysis.

Which LLMs have you used? The two you mentioned earlier, but not much. I've used the cheaper version of GPT to get a sense of what it does, but I was not overly impressed. However, it is still early days.

How useful do you find LLMs? They're still in development and far from being a finished product. The potential for them to become useful is high, and they could be helpful in handling small tasks. But for now, they are still in the beginning.

How could Large Language Models assist an M&A consultant in their work? Yes, as I mentioned earlier, I believe they could help in many different areas. I do not think of it as a single model but rather as many different ones. For example, I expect that our data pool, such as Gain.Pro, will eventually introduce an AI application that allows us to scan the entire database much faster. This would focus on identifying good acquisition candidates based on past or recent deals. The program might even add a predictive element. But this would be a completely different model than one that helps recognize patterns in audit files and supports future forecasts for projections. That is a different application because it requires different knowledge. On other fronts, such as market development and research, another model would be more suitable for broader knowledge scraping to develop a well-rounded subject. I see it functioning in this way in the short term, over the coming years. Time will tell how much further it can take us. That is how I see it. I compare it to the internet, when the internet first emerged, no one anticipated that we'd eventually want to see other people's vacation photos, which Facebook and Instagram have made quite popular. But that was not something we envisioned as possible with the internet in the 1980s or 1990s. Back then, it was more about wondering what the major players would want to share. That has changed significantly with the fact that now even a local hairdresser can build an influencer platform. The foreseeable future is mostly about time-saving, summarizing, a very useful tool across many applications.

Are there other repetitive tasks for an M&A consultant that could be replaced or assisted by a Large Language Model? Yes, financial analysis, not just internal and external market analysis and company analysis, but financial analysis is definitely an area where AI can identify patterns and work predictively. Writing tasks as well, of course. But beyond that, I think a lot of legal work could be assisted by AI, such as drafting contracts, where we currently start from scratch each time. Many shortcuts could be taken, but we'd still need an operator to check, similar to how a metal block is checked coming out of a CNC machine. In the past, someone would operate the machine manually, but now a robot does it. The operator still needs to set it up and conduct quality checks. I think a lawyer would ultimately work with AI in the same way, they would set the machine to produce a good report and challenge and adjust it when necessary. I also think AI will be increasingly applied in M&A over the coming years. But I believe many clients in our segment, SME plus, will not easily trust an AI tool to guide them in selling something that secures their pension or is the most significant business deal of their life. Regular expectation management and a standard frame of reference will still be necessary, and that is where our work will remain. Human contact will be the most important factor for many years to come, in my opinion.

Are there sections of the IM that could or could not be easily written by a Large Language Model?

Yes, I think market development is probably the easiest. Because that involves scraping a lot of data, both nationally and internationally, across Europe and worldwide, sometimes even regionally. That is something AI can handle more easily. However, a lot of internal analysis, such as understanding how a company is organized, which employees work in what ways, and identifying key players, is something a Large Language Model would struggle with. There is very little information available about these things. If it is based purely on data sources, sometimes public and sometimes less so, it is unlikely to produce good answers. You need broad questions that can be well-answered, otherwise, there is not enough information.

How much time do you expect an M&A consultant would need to verify the accuracy of the output from a Large Language Model? That is a bit of an open question. It also depends on what question you are asking. If it is something you deal with frequently and have sufficient knowledge of, you can quickly see whether the answer from the GPT or whatever AI model is accurate. If that is not the case, then you will need much more time to validate whether the result is logical. So, the combination could vary widely.

Appendix G – Interview C- Taurus expert.

My research is about the application of AI and primarily Large Language Models in Corporate Finance. My study specifically focuses on the internal and external aspects of a company's analysis for the IM. The financial analysis is excluded from my research. The internal analysis concerns the company itself, while the external analysis deals with the market.

Let me introduce myself. My name is Interviewee 2, a colleague of Quinten at Taurus Corporate Finance. I am one of the six partners and have been working here for two and a half years. I was a banker for ten years. First at Rabo, later at ING. I always dealt with large companies, mostly in the Northeastern part of the Netherlands, from Harlingen to Dinxperlo and from Amersfoort to the German border. I have mainly worked with the manufacturing industry and some business services, but the majority has been in manufacturing. I made a brief move to the business sector at Vierhouten Pallet Industry, where I worked as the right hand of the managing director. We were supposed to do a kind of buy-and-build strategy, strengthening the organization, and then heading towards an exit in five, six, or seven years. However, this was in the middle of the Covid pandemic, so it was very operational. You could not do many acquisitions in such extraordinary times. The role did not quite fit. Then some former colleagues from ING, Employee and Employee, started here at Taurus, and that is how I ended up here.

Have you also done acquisition transactions at ING? Yes, I have always done a lot of acquisition finance, both at Rabo and ING.

Did you also guide the transactions, or mainly the financing part? No, just the financing. I did not do corporate finance at the bank. It was the financing part. Once a Letter of Intent (LOI) is signed, the parties always go to the bank to obtain financing. That is what we always did.

How long have you been working at Taurus, and in what roles? I joined Taurus in January 2022. Before I started, we agreed on a path to becoming a partner if I met the criteria. But I started just as a corporate finance advisor. In January 2023, I became a manager of corporate finance, and in January 2024, I became a partner.

What type of M&A transactions do you typically advise on? What do you mean by "what type"?

Mainly purchase or sale processes. No, most, 9 out of 10, are sales-side transactions. That is what I enjoy the most. I only do the rest when it is strategically beneficial, like purchasing or financial guidance, because it distracts from focusing on sales-side. But sometimes it is useful to do that. If we are doing a purchase now, we know there will be a sale transaction eventually. It is good to be the M&A advisor now, so they come back when they want to sell. Then you have that bond and stay involved. You become the go-to party when they want to sell. And financing guidance, like debt advisory, is only for existing clients, and it needs to have a strategic angle. Or, we say it is 25 to 30 million in financing or more because then you can also charge a success fee. But, as I said, 9 out of 10 is sales-side M&A.

In terms of EBITDA and staff size, how large are the companies you advise? That varies. It is within the sweet spot, which has become a bit of a sweet gap by now. We are talking about family-owned businesses in the mid-market segment. But it is really lower mid-market to upper mid-market, everything between 1 and 30 million EBITDA. I actually do things between 1 and 30 million EBITDA, but my preference is large business, really starting from 3 to 5 million euros and up. I enjoy that more because the buyer group is more professional, there is more stakes involved, and you can put more of your expertise into it. So that is what I find interesting.

Okay, and in terms of staff size, how many employees do these companies typically have? It varies between 10 and 500, for example. Sometimes it is even more, but that is usually the range.

What is the size of the transactions for the companies you advise? It is really everything between, if you consider 1 to 30 million, but let's say, sometimes we have a platform with which we collectively handle transactions generally between 5 million and 200 million.

Which sectors do you typically advise companies in? Generally, I advise in business services and manufacturing, the industrials. Not all manufacturing, but the industrials. Those are my favorite sectors. I need to understand the business model. It is very important for your analysis to really understand the P x Q. I often have less affinity with IT. I understand it, but sometimes I do not believe in it as much, so it is harder to advise on. I've always done industry and have a strong preference for it. When you have a large trade business doing 100 million in revenue, you need to move 100 million worth of boxes. A company can achieve that actually quite fast. When you have 30 or 40 million in industrial revenue, you are already a significant company. All the world's challenges come together in industrial companies, scarcity of resources, scarcity of staff, the energy transition, the Internet of Things, industry 4.0, and they're already working on 5.0. There is significant capex obligations, so it is very capital-intensive, often a lot of competition, and global as well. So I find that interesting. IT is also nice, but it is a different business.

Since you said you prefer business services, Interviewee 1 mentioned in my interview with him that he enjoys that as well. Do you often collaborate on that? Yes, actually, because we have quite a few cases in business services, especially in staffing and contracting. We usually do the initial acquisition discussions together.

When it comes to the internal analysis of a company, how do you conduct your research? The internal analysis, well, we have our IRL (Information Request List) at the front, where we request a lot of information. So it is already quite data-driven. The internal analysis is partly qualitative, where we discuss the organization, the equity story, where the company stands, its legacy, and milestones. This is done by requesting information and also through Q&A sessions with the entrepreneur. The information is often very refined and not typically available in public sources. Large corporations usually have better reporting, even internally, but SMEs often do not. The information is often still within the entrepreneur or the family, so you need them to even get the story on paper. That is the internal analysis. Financially, if you have the annual figures and do your analysis, it already says a lot. That is something we are not focusing on in this study, but it is relevant for internal analysis. It is mostly Q&A sessions with the entrepreneur and management, so providing verbal information is crucial. It can all be on paper, but many entrepreneurs are better at expressing themselves verbally than writing things down. Some prefer writing things down, so the effect is the same with the information you get, but the approach is different.

Are there other sources you use for this research? For internal analysis, that is tricky. You always use everything available on the internet, databases, the company's website, anything you can google, news, going through everything you can find. But this has its limitations for the segment we serve. It is always a combination of public sources and what you learn from the company itself.

How much time does this internal analysis take in general? You would need to do some timekeeping to say for sure. It varies by company. In general, if you consider 30 hours for internal analysis, 40 hours for external analysis, and 40 to 50 hours for financial analysis, that is about 100 to 120 hours. You are likely to spend more time, if you are also including writing. Q&A sessions can quickly take 2 hours each. But you also need to analyze and absorb the information. Yes, but you are definitely spending, like with Project Delta, where there are 3 shareholders and a full Management team of 3 people, you are spending at least 12 to 14 hours on Q&A sessions. And then processing that could take 20 to 40 hours, I think.

For external analysis, how do you conduct your research? So first of all the Q&A sessions that we spoke about. But it really depends on the company that you advise. Ideally, you have a full commercial due diligence done by an external consultancy, like McKinsey, Bain, or BCG. But that is not always feasible, as they can cost half a million to a million. We sometimes work with former McKinsey consultants who started a company on their own or joined smaller collectives, which are usually more affordable. That is ideal because they map out everything, markets, growth opportunities, both in terms of business verticals and geography, product differentiation. They do real research, surveys, calls, etc. That is the best-case scenario. If that is not possible, the next best thing is a sector report, which you can purchase, but those also cost 5,000 to 10,000 euros per report. We do not do that often but probably should do that more. If that is not available, we use all available sector information. I often use Bain & Company's website. They have fantastic sector insights. But their information is very macro. Conceptually, I always try to break it down, from macro to subsector, to the specific company.

Also, considering competition and positioning, it is good for the macro perspective. I build my external analysis that way, gathering as much information as possible.

So it would be like an upside down pyramid, where you start broadly and then narrow down? Yes, because it allows you to create a complete picture. You can use economic bureaus from ING, Rabo, ABN, CBS for statistics, there is a lot to find. We use Gain.Pro to dig deep and uncover various correlations to deals and deal drivers. That way we can identify key deal drivers. That is how I build my external analysis, always combining information from various sources. Ideally, you would have commercial due diligence conducted, as that is the best option. The next step is to gather sector information, which is often very macro-level, covering entire regions or global trends. You need to constantly think about how this information relates to the specific situation you are analyzing. The advantage of this approach is that you can benchmark effectively. For example, you can compare the media trends with the performance of your company to see how it measures up. Benchmarking is great, but you need a lot of specific information to do it effectively. For example, in business services, where we have handled many cases, we can benchmark well because we have a lot of data. That is the big advantage of firms like EY, Deloitte, and Lincoln, as they have extensive portfolios across various sectors, allowing them to benchmark companies within subsectors easily. If you are talking about IT staffing in the Netherlands, for instance, we cannot do that kind of detailed benchmarking ourselves, aside from the general data available. They can do it on a post-level, which is impressive and something we cannot replicate. Then there are the reports and the Q&A sessions, where we discuss how they view the market. Industry organizations often provide valuable reports as well. So, the goal is to gather as much information as possible. A big part of our job is getting to know and fully understand a company within two months, so that is the approach for external analysis.

Does external analysis take more time than internal analysis? Generally, yes, because internal analysis can be straightforward, organizational structure, direct and management teams, a timeline of milestones. But external analysis is more important because it is linked to growth potential. It is built up the same way, starting broad and narrowing down, mapping out growth opportunities, supply chains, value chains, which requires a lot of information. But those are the steps we need to take to improve our IM further.

Yes, I agree. And what about the writing process for internal and external analysis? How long does that take? I do not do that anymore, so you should ask Employee 1, Employee 2, or Employee 3. Generally, I would say two hours per sheet you create, good content and conceptually solid. Others might be faster, but that is what I would estimate.

Have you had any experience with large language models? No.

Okay, so you have not used them yet? Well, ChatGPT is also a large language model? **Yes** I have a premium subscription, so I use it quite a bit. I often Chat with ChatGPT in the car, especially if I have not had English meetings for a while and feel rusty. I do business meetings with ChatGPT in the car to get back into the rhythm of speaking English. I sometimes put SPA's in it to see what it extracts, which works well, but you have to be good at controlling the output. It is still somewhat generic in what it can do. The human touch is still important to control the output. This is the third time AI is being touted as causing a revolution, and now it will actually happen. You will see a shift where people who used to do input tasks will move to output control. Previously, people were afraid it would lead to job loss, but every revolution has led to job changes. We currently face a staff shortage, so it is actually good because it allows people to focus on tasks that provide more customer value. The limitations so far are that you cannot request sector reports from it. It can interpret and summarize well, but you still need to control the output. I've looked into GRASP once with Employee.

Is that another model? Yes, it is for finance professionals. I have not tested it but have looked into it. I am curious what it could do for an organization like ours. I think we could do a lot with AI, but the question is, what can we do? And, of course, we cannot just put our clients' data into that model.

For Project Bridges, I asked ChatGPT to provide sources on why hydraulic engineering is a growth market in the Netherlands, using relevant sources like major banks, governments, or industry organizations. It gave me good reports, which we've used in the IM. That is great. Those were publicly accessible, but the reports I am talking about are not free. They cost 10,000 euros, with research bureaus behind them that need to be paid. There is a limit to how much you can use it. We are now working on Project LIBERO and ChatGPT provided a good top

10 of global premium manufacturers that we used in the IM. But would've found those companies myself.

But you found them faster, probably? Yes, but you still need to verify those companies because it is risky to blindly trust a suggestion.

If you ask ChatGPT for ten names, it will give you ten names, whether they exist or not. But I've sometimes put in an SPA before a meeting when I was not fully prepared and asked it to highlight the SPA's key points, risks for the sellers, and what needs to be addressed.

Was it good? Yes, it was fine. It works, but you need specific prompts. **If you just ask for what is important, you will get standard responses. Rubbish in, rubbish out, so people need training on how to give good instructions.** I think very few people in our company can do that well.

Employee made a comparison between large language models and CNC machines. Previously, people had to set up the machine themselves, but now it is more about controlling the product the machine produces.
Yes, but as more data becomes available and the model trains itself, we'll focus more on the output. Forty years ago, they predicted we'd only work ten hours a week by 2020 because we'd be so productive. But we still work a lot because we've filled that time with extra efficiency and other tasks. The question is, what will this revolution bring? There is still a human touch needed for understanding and positioning the case. People still buy and sell companies, maybe in a hundred years, an AI model will do it, but that is not the case yet. I am curious how refined the model will become in taking over work. But it is already revolutionary, so I do have experience with large language models. I've been using them from the start. Rabo does not allow it. My wife works there, and she cannot use it due to security issues.

Privacy, because you have to input information, right? Co-pilot is from Microsoft, right? **Yes.** And Gemini is from Google? **Yes, Gemini is owned by Google I think.** Yes, I think they use Gemini. Co-pilot runs on ChatGPT, so I think they use Gemini.

How do you think large language models can help M&A consultants with their work? As we discussed earlier, quick data analysis, fast summaries of large text chunks, and very directional insights, think about this, consider that. All repetitive tasks can largely be taken over by large language models. This will also apply to financial analysis, you can put a financial statement in and get key insights. But you still need to understand what it says, so the expertise remains the same, but it is applied differently. You will allocate your time differently.

That is similar to why you prefer working in industry and business services because you want to understand the business model. Yes, that is true. You will have more time for the good things, spending more on customer experience, service, and communication. You can serve your clients better.

Are there tasks for an M&A consultant that are repetitive and could be replaced or assisted by an LLM? Yes, in line with what we said earlier, you could likely have 80 to 90 percent of external analyses prepared, with just 20 percent left for tailoring to Taurus style and case-specific details. It would be perceived as custom work but still involves your touch. Internal analysis could be similar, though with companies like ours, where information often resides within the company, it is still a bit tricky to extract that.

What about design? **Graphic design? ChatGPT can make images, but they often have typos, which is strange given that it is designed to generate accurate text. It can make nice logos, but that is not useful for us.** No, it would give you more options for making a PowerPoint. But as far as I know, we do not have that capability. **Recently, I used it for Bridges. I was working on an IM and asked it to create a color palette for a pie chart based on Taurus colors and the colors used by the client. It provided color codes which I could use directly in PowerPoint.**

How much time do you expect it would take to review and improve text generated by a large language model to meet Taurus's quality standards? That is an abstract question. It depends on the output. I cannot say right now. The question is clear, but the answer is abstract. A single page could be done quickly, but if you receive 15 pages of output, it will take more time. Could you try rephrasing this question? This is a vague answer, but it is true, it depends. What do you expect?

Say it gives a two-page text. You need to rework it into Taurus-style text. How long would that take? If you get two pages of input from the LLM, you can finish it quickly. But the big question is, do you know if anything important is missing? **Yes, that is the big question.**

But if you think it is sufficient, you can move on. It could save a significant amount of time, at least 50 percent, as you will only be focusing on formatting only.

I said that the time savings are mainly in the writing part because you still need to do the research beforehand, regardless of what ChatGPT provides. Yes, research is necessary, but some tasks can be handled by an LLM. I think you could save between 40 and 70 percent of your time over the year. You could then spend that time on design, storyline, etc.

Appendix H – Interview AI Expert D

My name is Quinten Kropmans, I live in Oldenzaal, and I am studying Business Administration at the University of Twente. My master's thesis are about the application of artificial intelligence in the field of Corporate Finance and M&A. The research specifically focuses on the internal and external analysis of the companies we do business with. I would like to interview you about what you can tell me about this topic.

Great, that sounds good.

Could you first tell me a little about yourself?

I am [NAME], an entrepreneur since the age of eighteen. I've always been in technology, and my background is also in technology. I am a technician by origin, and in the past, we only developed custom software. We operated around the world, with offices in the Netherlands, Miami, and South Africa. After that, I started focusing on a different direction to have a smaller custom-made branch and eventually use AI to manage customer contact more intelligently. So we have experience with that in a custom-made branch and do a lot with it. Ultimately, in customer contact with our own products and developments, entirely our own. So, currently, I am only generally responsible for the club, no longer for the technology.

Which companies have you worked for in the past?

I have only had my own businesses. Okay. So that is what I've been working on.

Did you always have [Company Name]?

No, I started with [Company Name]. So, I started with [Company Name], I sold sunglasses online, eyewear, I sold paint, I sold jewelry online, invested in a knowledge platform, Enlisted online. I also sold that. I had my own advertising agency, which I also sold. Yes, that. I believe they call that a serial entrepreneur. So, it is all somewhat related to the chain, something that strengthens the business. Currently, I am solely focused on IT.

What is your experience with large language models?

We fully utilize them in our applications.

What do you use them for with your clients?

We use them to index documents. From there, we train a language model. From that point, we determine how to collaborate with language back and forth. You can train an LLM, you can handle it. You can model it to create content for various applications.

Which large language models do you often use for that?

Which ones do I often use? Quite a few, actually. We even use the older Gemini 1.0 for that. We use the LLM Gemini Fast. We also use Aurora, but that one is entirely private and needs to be implemented. We have done that.

So, is it the case that one large language model works better for one application than another large language model?

Yes, definitely. If you look at the popular LLMs, for example, you have ChatGPT, which is, of course, more than just an LLM. But there are different models. You also have models that are better with language, in Dutch, in English. So, you can use those better for interpretation. Gemini has made a huge leap in that, as I mentioned to you. There are models that are better at calculations. So, to extract numbers from the model and then calculate with them. So, it really depends on where the focus of your language is. Is it in interpretation? Is it in the next steps to do something with it? The models do differ in that.

How useful do you find large language models?

Very useful. In what sense are you asking?

I also asked this question at Taurus. Generally, I got the answer that they've used it once. They've used ChatGPT. I wanted to see what the difference was between how colleagues at Taurus use it on the work floor and what an expert would use. I was curious to see how much difference there would be. And how useful many large language models are. They said, "Yes, I find it useful because it can rewrite text." But as an expert, you would mention completely different things.

Yes, look, the nice thing is, of course, you have different models. Let's take two, for example, if you take ChatGPT now and Gemini 1.5. Gemini 1.5 is a bit faster. On average, a request takes 2.5 seconds. So, if you want direct interaction, it is not so great. If you want to rewrite texts, it can be quite okay. However, a language model is just a very good predictor and does not always understand the context. It is simply a data magician that you can teach. But understanding context is very difficult for it. So, especially when it comes to marketing texts and rewriting existing texts, yes, it is really perfect for that, and I think it is also perfect for completing texts. You can use it very well for that. So, I think if you are uninspired, thinking, "How in the world am I going to do this?" Yes, you can use a standard LLM very well for that. They are excellent text generators. Yes, and most of the time, the challenge lies in coming up with texts, how to put them on paper, how to phrase them correctly. Yes, and when you are uninspired and do not know how broad you should go. You can give a nice prompt to an LLM, and it can generate a nice little text. Look, if you want to create an analysis for a company or make a nice introduction text for a company, or you want a backup procedure, it has all kinds of existing prompt texts that you can extract. So, in the open structure, you can get quite a lot out of it. But if you really want to implement it in your company, then you need such company-specific knowledge. That is where you encounter a different aspect, which a standard LLM does not handle so well. Because, in the end, it is just a prediction mechanism, so it includes a lot of data. But most of it is standard data, and what makes the difference, that is not always as refined in the current LLMs.

Do you always need to verify the information that a large language model gives you because it can also contain misinformation? Yes, look, it is a language magician. What people often do not realize is that it does not truly understand language. It is just an algorithm with language. Suppose you ask it a question, something like, "Can I get an exemption? Can I get a parking exemption?" And the suggestion an LLM makes, writing a text about that, is "I have a text about a parking permit." Yes, they are somewhat related, and in a municipal context, they are somewhat connected, but there is actually a difference from what you as the writer intend. You need to be aware of that. These are small things that you can easily overlook. So, there is still a challenge in making sure you verify it. Also, in rewriting or ensuring that the context and the emphasis you place are correct. Yes, you definitely need to do that. I must say, that does take quite a bit of time. Because look, you are operating at that level. It is, of course, amazing.

Just checking and screening to make sure there is nothing strange in it, that is 10% of the work, compared to 90% of your work just to get a text on paper in the first place.

Could you say that a large language model is actually a kind of calculator that calculates, based on your prompt, which words should come out? Yes, pretty much.

Can you clearly notice differences between the various large language models? That the application is actually quite different depending on the large language model. Yes, how can you achieve that if you use, for example, Gemini? It can be entirely based on current data. So, you can use it for more short-term searches, but it is always lighter. But if you use ChatGPT, for instance, it has data up to 2022, if I recall correctly. They've worked on the texts and the context they've organized. They, of course, have scraped all of Wikipedia. So, depending on what you are going to use texts for or what you want to generate texts for, the sophistication of the "calculator" matters. The second is what data you are allowed to use. So again, if you are working within your own data cabinet, you can train it separately. In that case, some other models or a combination of models might be better. So yes, you can definitely work with that.

Is it clear in advance which model you should use in which situation? Most people do not get further than just turning on ChatGPT or Copilot, and you can simply play around with that. But look at your application. Take Plaud, for example. If you use Plaud, then you have a very clear picture. Namely, you just want to save texts and listen to them, and there needs to be a good transcribing model, followed by a few actions. This is not a very exciting application. So, you are just looking for a cheap model that can do that. OpenAI has a great API available that does this, and Plaud smartly uses it. So, you record an audio file, put it in a tool, and then add different processing models that ChatGPT can handle. Yes, so the combination is smart. Is ChatGPT the right tool for that? Well, for you and me, for the Dutch language, the functionality with Gemini would currently yield better results. Gemini is more advanced in the Dutch language than ChatGPT. You will notice that. If you look at the text, how it is translated, and how it then processes and performs actions. Sometimes with ChatGPT, you need to rework the text a bit more than with Gemini. Yes, give it a few months, the development is happening so incredibly fast that you also have to wonder how long it will be before ChatGPT catches up again. There is simply a major power struggle going on. So, it is crucial to determine in advance which model you want to use for what purpose. For example, if you are just generating texts for reports, you have a very different need than when you are using the models for data analysis. If there is an AI engine like GPT-4 behind it, then the data is up to 2022. That is not very helpful now that we are in 2024. So, you have to realize that it is important to know what your goal is and what you want to do with your data, especially also considering what your sources are.

If you need to take privacy into account, for example, when we work with companies, all the information they share with us is private. This information is only shared with other companies when they have signed an NDA. How can we take this into account when using large language models?

Yes, what you can do, and this ties back to what I mentioned earlier, is if you use Gemini, you can load a model locally. With ChatGPT, you can also run it in a secure environment. So, you need to set up your own secure environment. First, determine what the purpose is. So, why would you want to use AI, and what would you want to do with a language model? For example, do you want to create a report, and therefore need the ability to interpret language? If you have a lot of Dutch language or just language in general, then Gemini is currently the better option. You would need to create your own "cube," so to speak, your own data cube where you can secure that data. There is a way to run it on your own server, and you can check a box to control whether the queries you make to an external LLM are allowed to store or use that data. This is where the big battle is happening now because how do these models learn? OpenAI has not been very transparent in the past about where they've sourced their data from. If you are very transparent, then you have nothing to hide. So, this has been a hot topic in the media, how do we secure this properly? Yes, you can check a box. So, if Taurus wants to use OpenAI, you can do so in a secure way by using OpenAI's cloud, checking the right box, and then assuming that box is interpreted correctly. But you have to ask yourself, how clear is that box? Is it enough? The same goes for an NDA. How clear is it? You generally assume that when you sign an NDA, the information stays within the agreed boundaries and among the people covered by the NDA. But you never know if someone shares information at a party or at home with a partner, something that is not in the NDA but still happens. And you may never find out. But it is kind of the same with an LLM here. You just do not know. You can check that box, but you do not have a 100% guarantee. I think you have to approach this very pragmatically. You need to know what you are putting out there. So, one option is to close it off within your own environment. In your development process, like what Goldfire is already doing, you put it in your own environment. You can do this with OpenAI and with Gemini. In your own environment, you can train different applications on it. That is one thing. You can also use public LLMs and then indicate that the data you share must not be public. And in both cases, you have to assume that it is technically well implemented.

What are the capabilities of large language models that can be useful for an M&A consultant? Our tasks are mainly focused on M&A, such as writing an IM (Information Memorandum) or a pitch before starting a project. How could language models help with these tasks, and in which areas?
Well, as we briefly discussed in a short meeting, in your case, if you have all the files, a lot of the work involves analyzing the files you receive. The great thing is that AI can analyze this much faster. With a good language model, you can ensure that it also draws the right conclusions. So, if you have a secure environment where you can input all the data from a client, you can use that data to run a number of prompts to create a report. Based on the data and quotes, you can use this to do something meaningful. I believe that is the first and second thing.

What you can also do is conduct a market analysis based on the data you have, compared to the data and competitive analysis available in the market. You can let a model handle that and have it write in the language you want. You can set this up very quickly, making significant progress in a short amount of time. It can go as far as creating a practical case. There are simply tools where you can say, "I want to create a website, this is the segment." If you want to make a website for sports articles, you can just give a standalone task point. If you want to analyze the market, see what the most popular sports articles are right now, then find a good supplier with margins that can deliver with drop shipping. If you want to create a website that integrates these two things, that works together, and if you want to set up payment profiles, you can do that too. Then, you can create a Google Ads campaign to offer all your products well and integrate it too. You can create a research point to have Google Ads written for your web concept by AI. You can perform analyses, generate texts, and the combination of these things allows you to analyze a web shop very quickly in this case.

So, in my opinion, if you are an M&A consultant, this is what it is all about. You want to quickly assess the market analysis, what the company's own concept is, what the market is demanding, how to align that with your current product, what the strengths and weaknesses are, how you would write a marketing plan, and what the opportunities are. You want to analyze all of this quickly. If you do this with a format, and if you look at what an M&A consultant does, in my definition, it aligns with the example I just gave. Essentially analyzing a web shop, but on a smaller scale. Ultimately, you want to analyze the market of the company itself, the potential market in the world, in the city, or in the region, depending on how broadly you define it. Then, you look at the possibilities, what the competition is doing, and what steps you can take yourself. So, the entire analysis process consists of several steps that an M&A consultant goes through. Normally, you do this analysis in your head, but you can certainly use AI for it. An M&A model is trained on data, of course, which you are still researching. So, I believe that you can have 90% of your work done very well, but you will need to add or correct 10 to 20% yourself to make the difference. But you will save yourself from doing that bulk of the work.

Do you expect that external analysis might be easier than internal analysis of a company, for example?

Well, I do not think it makes a difference. If you just set up the data, you can do it. The thing is, neither analysis is accurate. The external analysis never has up-to-date data because it has not been released yet. So, you are always working with data up to 2022 or 2021. Look, for example, at economic conditions. In two years, a lot can change in an external analysis. So yes, you can get quite far, but up-to-date information, right up to today, will not be completely accurate.

The same goes for internal analysis. Before it becomes a document you can analyze, it takes time, both internally and externally. So, I think it would be a waste not to use an LLM for your analysis because anyone can do it. Let's be honest, it is available technology. If you do not do it, someone else will. You also do not know where you can be distinctive, so that becomes more challenging. The key lies in the information that has not been processed. Documents with text have been processed. Not all information that is known is directly shared with models like ChatGPT or Gemini. This is true internally as well. You might find something about a company's performance if a financial statement has been filed, but that financial statement can be delayed by a year. So, that is not very helpful. You need to have the data, which is, of course, crucial both externally and internally. And internally, you do not know how processes work. Processes and the interrelation of things are difficult to analyze. So, it is not like you can immediately say, "Hey, I have my ISO certificates, so the business processes are running smoothly." No, if there is a lot of staff turnover and the processes and quality are declining, then you can put that into a model. But if you look at the atmosphere or culture in a company, which you also need to consider, you need to talk to people and observe to identify pitfalls. Is someone always optimistic? Are you dealing with a company salesperson who's always positive and views the numbers differently? Or is it the opposite? In a company, you also have to deal with a bit of culture and processes and things that are organized but not shared, which you, as an M&A consultant, of course, need to uncover.

So, you can only analyze once you have the data, and "measuring is knowing." But you need something to measure. That is the danger, I think. Because you can produce results so quickly with a good language model, critical thinking might be forgotten. Anyone can use a language model. I think that if you use it now, you are a pioneer initially. But when you see how many companies start using it, that will not take long. So, the question then becomes, where am I still distinctive? It lies in what you cannot automate. How do you, as a company, add value in that area? That is where the strength is, and how you can distinct yourself. So, indeed, it is about helping with drafting a text. You still need to add the final touches, correct, and improve it yourself. And particularly, you need to ensure you have all the data. How reliable is the data I have? So, you are not really being distinctive as a company or individual if you just throw everything into ChatGPT. You are not standing out anymore. So yes, if you and I use the same prompt, there is a very high chance we'll get the same output. It might just be worded a bit differently.

What is the best way to prompt a large language model to guide it toward generating a useful text?

Yes, I find that to be a very interesting question. It is a very interesting question. What is the best way to drive a car? Look, the best way to prompt is to ask ChatGPT itself. And then it will give you different possibilities for getting the best result. It involves a lot of trial and error. So, before you figure out what you get, you will first want to develop a prompt yourself. A prompt is nothing more than just an input query. It is just, "What am I asking? How do I create a good prompt?" If you want to do that and ask something with it, you can type that into a field, but again, that will not get you very far. I think the first step in figuring out a good prompt is to ask questions and then review the results and fine-tune how to get that good result. You can also search extensively for prompts. There are plenty of websites that have already released many prompts, along with the lessons learned. You can easily use those. As an end user, if you want to start something from scratch, then it is about putting something in and getting feedback, and then asking again, "How do I get a better result?" These LLM models are based on an algorithm. They make a prediction of the best word that comes next, and how and why. It is just a mathematical formula, so if you expect a different result, you can simply ask the LLM model. You can ask, "Can you make it faster? Can you make it more efficient? Can you approach this differently?" Eventually, you will figure out, "Yes, now I am getting a good result." Then you will look at how you can formulate your prompt as quickly and efficiently as possible. But first, you want to know what comes out. So, it is a very tricky question you are asking because there is not one way to say, "What is the ideal prompt?" It is a very short, clear, efficient formulation of your goal. That is the ideal prompt. And that can change as the LLM model evolves. A good prompt in version 3.5 or 4.0 of ChatGPT produces different results. Those results are also different if you use them in Gemini.

So, it also indeed depends on which LLM model you use?

Yes, definitely. If you say, "I want a general text in an informal or formal tone," it is not too complicated. But if we say, "I want ten synonyms for a specific word or other client queries," then it becomes a bit trickier. If you ask for one sentence to be rewritten with an additional alternative sentence, to give you an idea, if you want five alternatives, it takes two and a half seconds. If you want thirty alternatives, it takes nine seconds. Now, if you want to work efficiently and you want to perform this action for three thousand sentences, I do not need to explain to you how long that prompt will take. Some people will not be happy with that. But the fact is, how good is the language model, and how good is your dataset? Are five results enough for you in the example I gave, to create synonymous sentences? Or do you actually need thirty sentences? If you actually need thirty sentences, then you need to consider whether the prompt you have given is the most practical for generating such a large amount. So, you also need to consider how well the model performs and how it does that. Based on that, you might need a long prompt, or a series of follow-up prompts to achieve that. In the end, it will process that very quickly too. So, there is not an ideal prompting method. The only thing I can say about that, in response to the general question you asked, is that it should be as all-encompassing and efficient as possible.

I've noted in my research that finding the right prompt is probably best done through trial and error. Because it does not take much time, of course, if you just try it out. The model generates within a certain timeframe, so you can wait for that.

That is correct, but to go back to your original point, it does that in just a few seconds. That is ideal, but the challenge is also how to validate a large dataset. How does it do that? People sometimes underestimate this. Look, this might be a trivial example. If you ask for a sentence from one A4 page and ask it to analyze and summarize that, it does that well. But if you throw 3,000 A4 pages at it, how do you know it is considering all 3,000? How have you validated that in your prompt? You can scan that one A4 page as a human to see what is in it. But we cannot scan 3,000 A4 pages. So, if you throw 3,000 A4 pages into the LLM in your prompt, and the prompt thinks, "This is going well, I'll have something for you in a couple of minutes," how do you know the prompt did not stop at page 100?

How do you figure that out, by doing research yourself?

Well, you need to take that into account in your prompt. It is helpful, and experience plays a role, but you need to be aware of it. It is not like you automatically know your formulation is correct when you create a prompt for something small. Then you move on to the next step: how scalable is it, and how do I validate all my data? Eventually, you could say to the LLM, "As you have written it in the summary, that is great, but could you also tell me at the end of your conclusion how many words you used to analyze?" Now, if you have this document in Word with 3,000 pages, you can see that it contains 10,000 or 20,000 words. You can see that in the document's properties. So, if it says, "I used 2,000 words," you might think that is odd because that is too few. So, you also need to consider the integrity of your analysis. That is often forgotten when creating a prompt. How do you then validate your prompt? How do you validate that the LLM did not encounter a connection problem halfway through? You hear very often that when you engage an LLM, it sometimes just stops generating halfway through.

Beyond what we just discussed, like validating your answer, preventing the spread of misinformation, or dealing with privacy, what are the main dangers of using large language models?

Yes, I think one of the main dangers is that the user might become complacent, and you may not have a clear way to verify what is actually true.

So, basically, you might automatically interpret the LLM's answer as the truth?

Yes, we literally see this happening in practice. Look, we are used to looking at Google, and Google has 90% of the truth. But if ChatGPT, Gemini, or another model does this, and there is an error, we tend to also perceive that as the truth. So, if something happens in an analysis and you ask an LLM to work its magic with its data, we assume it is doing it correctly. This increases dependence. If you are 90% correct, you probably will not doubt the remaining 10%, so you will not question it. If you have a good recipe for potato pasta or pizza dough, and an LLM generates it neatly for you in an easy-to-follow step-by-step plan, and you do not understand it, you might ask, "Can you write it out for a three-year-old?" Then you get simple steps so that you can bake something, and that is great. You will not question whether it is correct or not. With the same ease, you might throw in an analysis of an entire company. Since the LLM has always helped you so well, as a human, you automatically start programming these patterns into yourself. So, if the LLM has always proven to be a reliable source for you, the danger is that you stop using your common sense and do not think critically, like I mentioned earlier about analyzing those three thousand pages.

That you just become too lazy as a person.

Yes, you become too lazy because the speed is so great that you also need to think about how to validate that high speed with the data you are working with. You need to have a policy for that or do something with it. That is definitely dangerous, yes.

What are the expectations for the future of large language models?

What do you mean by expectations?

Well, the developments are happening very rapidly. But where do you think it is headed in the future? Do you think AI will completely take over, that certain professions will inevitably be taken over by AI, or that tasks will be completely transformed? What is your perspective on this?

Well, I think that generative AI will be the biggest change since the industrial revolution. I think back in the day, people used horses and wagons, and then the diesel tractor came along. We all thought, "Wow, that is great," and it happened so quickly that eventually, horses and wagons became obsolete. Now we only work with diesel tractors that can cultivate entire fields and achieve massive productivity. That means that many people who worked with horses, blacksmiths for horseshoes, horse trainers, horse breeders, horse caretakers, no longer have jobs. We do not really think about it, but that was less than a hundred years ago. If you look at the period after World War II, there was an enormous technological progress. In fact, with computers and automation, much of the low-skilled work has been automated. Manual labor has been automated, and then we automated office work.

What is happening now? In the past, you had to learn. If you wanted to be a lawyer, you had to study. You had to know a lot of books by heart, legal texts, and so on. If you are an experienced professional, like a director or a manager, you have learned a lot through practice. AI does the same thing, it is learned everything, but from all the companies combined, and it gives you answers in just a few split seconds. So, I think the impact will be huge. Why? Because, as we just said, what does an M&A manager do? What could it mean for HR? AI can be used for writing HR texts, it can handle processes. So, it will have an impact on automation, on marketing, on sales. I think you will need half as many people, and you will need very different kinds of people. But we have a labor shortage, so there will still be a need if you want to grow. But yes, I think the impact will be enormous.

Yes, a colleague of mine compared it to the following example. In the past, you had a mechanic who would make a part or something similar. Now, the machine makes that part, and the mechanic has to check it. You can compare it in that way.

Yes, that is true, but I think many companies will see many functions in their current form simply disappear. There are a lot of "dumb" functions, although it is not very polite to say it like that. But you have a lot of those postman-like roles in business, especially as companies get bigger, with all the excess baggage that comes with it. It is just not efficient. The number of meetings and discussions you have to align things is immense. So, productivity becomes much shorter, much faster. Therefore, the purpose of the work also changes. You have probably experienced in your studies that you have different teams of people, like with the Belbin tests. You have different types of people. One is very directive, another wants to discuss everything extensively. That is all necessary to reach a conclusion or to work together. We as humans need a certain composition, but AI does not. AI just provides a lot of data, and then you have to validate it. That requires a lot of skills from one type of person, which is validation. But creating, like an inventor, coming up with something innovative, becomes much easier. I think it will be a big problem for people if they lose job satisfaction in certain roles. Especially in creating, thinking, and collecting, I think AI will have a huge impact. You already see that happening, it is not for nothing that people in Hollywood are striking as scriptwriters.

That they're basically being replaced, you mean?

Yes, they're just being replaced. You can just have AI write a nice script, right? Analyze the top 10 best films that won a Golden Globe Award and create a good script from that. The same goes for marketing texts, look at advertising agencies. If you are an advertising agency and you still have to come up with ideas. Suppose you just graduated, you are 25 years old, you have just finished school, and you need to start writing marketing texts. You'd better find another job. I think you just need to learn how to use a large language model and make sure you can work with it effectively so that you can be productive. But being distinctive is super challenging.

Yes, I agree. I think the same way about it.

Right? I think it is really something. But look at a teacher, what do you still need to teach? We are programmed to learn something. So, you are 25, you have just finished your degree, had a nice job, done your internship, studied Dutch, had marketing lessons, history lessons, economics, to understand everything. You have learned all kinds of facts and done assignments, and in the end, what do you do? You enter a prompt into a little program. Great, that is something the teacher would have done. It is really bizarre, so we are moving toward a system. It is a long answer to your question, but I really think that for the next 10 to 20 years, especially for people your age, you have experienced both. I went through that when I was 18 when the internet came along. I think it is an amazing time. You adapt to it, I still remember when there was not a computer. You are still part of the generation that remembers when not everything was AI-driven. You are now working with older colleagues at Taurus who still prefer to do things themselves. They think, "What is the impact? What does it mean for the future?" They only have 10 years left until they're 60 or 65 and they retire. But if you want to be innovative and ready for a new job in the future, yes, you will have to adapt. So, you look up the 10 jobs that will be replaced. I believe that includes programmers, marketing people, customer service employees, financial analysts, accountants. Everything in the financial sector, with analysis and tasks, is already there. So yes, it is a big change for the world.

Agreed. That change is indeed happening very quickly.

Yes, I am not saying whether it is good or bad, but a lot is going to change.

This was the last question of my interview.

Great, man. What do you think will change? For the future? Because you are doing your research for a reason.
I think the change is happening really fast. Maybe too fast to keep up with, because there are so many different applications. Yes, I think jobs will either disappear or change drastically. And I think, for example, that an accountancy firm might be the quickest to be affected. Or that one person might eventually be able to do the work of a hundred accountants now, simply by running everything through ChatGPT and just needing one signature. Yes, they can, based on what GPT says, determine if it is been done correctly, if all considerations have been taken into account, so you can eventually see that. It is done the same thing another accountant would have done, but in fifteen minutes, for example. So, entire accounting firms could become much smaller, in terms of the number of people.

In terms of the number of people, yes, I think you are seeing it clearly.

What did you think of the interview? Did you find it a good or interesting interview?

Yes, I thought it was a fun interview, definitely. From your research, I expected more depth in the questions, particularly around data validation. Because I think that is currently the most challenging aspect. We talked about it briefly, as you mentioned, with a good prompt and such. But also looking at what is happening in the market. You had already given some examples, but you had not incorporated these into your interview beforehand. So, it would actually be interesting to ask people, "What tools do you use, and how do you validate your data?" Because there is so much potential, you also need to get a sense of where to start. And I have the impression now, based on the questions you asked, if I am being critical, that you do not know where to begin.

Where would you say I should start, then?

Well, that depends on your objectives. Look, you need to research something, of course. So, I do not know what your project's objective is, but it is very important, no matter which direction you choose or what you want to research, to have a clear goal. AI is a very broad, overarching concept. If you want to set an objective within that or give advice toward an objective, it is helpful to be able to ask more concrete questions based on that. Because this is indeed about where to start and how to lay out a good analysis. I think a lot of analysis is going to change. Okay, but as I just answered, one model is better than another for analyzing data. So, how do you connect those models? How do you validate those models? Are there already developments in accountancy where you can analyze so quickly? Where you can create a financial statement, as you mentioned, and analyze it? And what data do I generate for my market research? How quickly is that already happening? And which parties are already doing it? What lessons have been learned? You can already see that online. What is your goal? Do you want to start working on this by the end of the year? We know that the pace of change is fast. That means you have to keep up with it as a person. I think you are right that this is like a snowball effect. At some point, all the neighbors have a diesel tractor. You are still figuring things out, thinking, "I just got a new horse, and I still kind of like it." But everyone in the country already has a diesel tractor, and only one person has a new horse, and that is you. Before you can drive your diesel tractor, you need to gain some experience. The first time, you might drive into a ditch. Then, you will fiddle around with the mower a bit. It also takes some time before you are sitting on that tractor thinking, "Well, this is actually pretty good." But then you realize you need a different tractor. That realization that you need a different tractor only comes after you have used it. You have been on a tractor. Darn, this is not a convenient tractor. Maybe that tractor cost half a ton, but now you realize you need one for 30,000 euros or one for 100,000 euros. But at least you have gained experience. That is the price of learning, and it applies to both companies and individuals. So, your goal is important. So, really, start as soon as possible to gain experience and see, "Yes, I want to work with this robot. I want to get a feel for it." Because not all technology is pleasant. When you buy your first car, you have different needs than when you are sixty and buying a car. Both are cars with four wheels and a steering wheel, but you have different goals. When you are young, your goal is to have a cool look, to make an impression on yourself. The car should be a bit fast too. When you are eighty, you probably do not need to impress yourself anymore. If I can park safely, I am happy with that. So, you have a different goal, but you have learned that over time. You can make the same difference here. I think it is also useful not to look at it theoretically, but at the practical building blocks. So that you can eventually take a step forward in an organization.

That is very difficult for a company because it is so broad, big, and all-encompassing. You need to gain experience with small steps. I think it is really important to incorporate that. What steps can you take? How do you ensure you get "tractor driving lessons"? How do you ensure that you get that training? You will not learn it anywhere else. This is the only area where you will need to do something with it. As you said, it is trial and error. So, ultimately, when implementing processes in your company, you need to work on a cultural shift. To use another example: I have a horse, and it just keeps working and plowing my land. But I also have a tractor standing there, and I need to use that tractor as well. If I take that tractor out on the land when it is raining heavily, I'll just sink into the ground with my wheels. You realize that the first time it happens. That is the same with your first prompt, you use it, deploy it for report analysis. You have bought a tractor for 50,000 euros. Then you think, "Well, great, let's go, it is nice and dry, because I get wet on my horse." You drive the tractor across the land, and the whole field is ruined. Well, that is a bummer, and you are three times as busy fixing it. Yes, you paid the price of learning. You realize, "Okay, if it is wet, I should not take that heavy thing onto the field." You make that mistake once, and then you know how wet the ground can be before you can drive on it. That is trial and error, while if you have older folks in the company, they already know this. They figured it out with horse and cart. They know, "I can get across with a cart, it'll be fine." But with a tractor in a turn, you will tear up the whole field. That is a practical example, and it applies to this development too. You can wait until you have completely mastered the tractor, but no one's going to teach you how to use it. You already know it because you see this development happening so quickly, new tractors are coming out every day. But several companies are not even using those tractors yet. But new tractors are coming out every day. So, to get a sense of those new tractors, you need to have been on a tractor. You need to get started and accept that you need to take driving lessons, that you might buy the wrong tractor, and that you might ruin the field once. But you will also think, "In September, I need to harvest the corn. I'll be done in a day." That used to take a week, so that is great. That gives you the energy to take the next step. So, learning, this question of how to implement it and what the possibilities are, is also very much about a cultural change. People in the company are just busy, and that is what I meant by the example. You are busy plowing the field with horse and cart, and that continues. But meanwhile, there are also new tractors ready to go. How do you ensure you have the time in your organization to free up money, resources, and people to occasionally play with that tractor? To set aside a piece of land for that tractor. So, the store keeps going with horse and cart, and the rest with the tractor takes a look at how it is working. You need to learn, stand up, and observe, that is very important for a company.

Because what you are saying, I'll repeat it, is that it is trial and error. That does not just apply to the technology but even more to the people. The younger you are, the more flexible you are, and the older you get, the harder it is to adapt. It is also about letting go of experience and validating it. But in this area, you really need everyone and everything to address all aspects. How do you validate a good report? You and I cannot take a report from someone with the experience of a sixty-year-old. If you look at someone like Assessor 1 within Taurus, with all the experience he has, you are not going to achieve that with just an AI prompt, it is not possible. But I do not see Assessor 1 creating a prompt that is decently fast either. I think we could get pretty far in an afternoon. You'd have a prompt running, a report generated, and you'd get some results, which is cool. Then you could ask Assessor 1 to validate how good it actually is. Maybe the first time, he'd say, "Well, I am not impressed." But eventually, it might turn into, "Look, a few hours of work, how long would it take you to do this?" That is how you build toward adopting it, and it takes time. As you have already noticed, sometimes you are so busy at Taurus that even sending a reply takes time. Let alone developing a prompt, let alone developing with AI. So, the tip and the thought I want to leave with you are this: If you want to be ahead of the curve and implement this, it is not just the technology that needs time. Often, when something happens quickly, there is a tendency to sit back and watch it pass by. But the awareness and the research, to bring that to everyone on the team, to take small steps, that is crucial.

It is like the tractor comparison again. You can stand back and watch it from a distance, but if you start too late, you are not going to master that tractor very well just by sitting on it once. It means taking a half-hour every day to drive it. You asked a good question, and I am giving you a very detailed answer. Even in your research, this is something to consider in your project. Where do you start? That is really challenging because it is so vast. That is something to keep in mind. Let me put it this way: What would be very applicable for Taurus is to take steps and do small things. Since you are doing the research within Taurus, I think it is a great thought process for you, starting from a light-hearted research idea, but then figuring out how to define your milestones smartly. How do you move from a larger, strategic plan, yes, AI is really happening in the future, how do you translate that into a scope for the next three years and a scope for today? How do you make that happen? And I think that is something important to incorporate because that is where the challenge lies in such a change.

We've decided to propose to Employee that we want to use ChatGPT and Gemini for each employee, maybe with shared accounts. But right now, the process is taking so long. Employee has to discuss it with various people, and then it needs to be rolled out. So, there is quite a bit of noise in the process, which means we cannot provide feedback as quickly. So, as you said, the time is not really being made available yet.

I come across this often, so I find it interesting to mention. It is a flow that happens in every company. You need to think about how to ensure progress continues. The "tractor" keeps moving forward. That is why I like sharing this with you, because if you look at how quickly discoveries are being made, there is a new tractor on the doorstep every week. But when a new model comes out, it is not so exciting. A new LLM solution or application to play with. And that is kind of strange. It is happening incredibly fast, but maybe too fast. That is also the danger, you need to determine how you are going to manage it. How do you ensure that development continues? How do you secure each step? And when do you escalate? "Hey, I am already 10 models behind." You are expecting a learning curve from me, but you cannot go any faster. The team cannot go any faster, so you have to keep moving through the learning curve to pick it up. We applied our first LLM 2.5 years ago. We taught ourselves machine learning, so if you look at the basics of how things work, the thought process you go through as a team or company is very different. Even the skeptics, I know with certain queries you definitely should not do it, because it is too slow. So, you cannot apply it in real-time, you have to work around it. You discover that over time. But because you know this is a tractor that can go 80 kilometers per hour or this is one that can go 5 kilometers per hour, there is a difference. But if you only need to plow the field, you do not need to go 80 km/h across the land. But that experience comes from doing, from trying. How do you secure this progress? And also, how do you escalate? Who's responsible for ensuring this, Employee or you? I am not talking about people, I am talking about the project and the learning curves needed for these steps. And you are going to need a certain number of hours in the learning curve to get somewhere and to experience things because it takes time. You say it is not difficult and it does not take much time, but it does. Not to type the prompt, but to experience the prompt and to process in your mind, "What does this actually do?" That is also a very important aspect. Do you want to make this change continuous? Because it is really something you have to go through.

Driving a car, you take lessons for that. On average, it takes 30 driving lessons before you can learn to drive. Yes, you will need to spend 30 hours on that. You can do it in 10 days, but you can also do it in 10 weeks. Yes, that is a good example. It is a simple example, and it applies here too. We are basically going to take driving lessons. We say it is not difficult, and we need to do it, but we know we need lessons. The lessons are a bit boring, but we'll take one lesson per year. Well, that is the challenge between innovation with AI. If I want to do something with it, I'll need to make time. I need to learn to drive. It is a huge turning point that you need to start taking lessons. In this case, you do not have a teacher. You can approach companies like ours to take steps. But it is also about getting behind the wheel yourself and sometimes crashing into a wall or driving into a ditch. Then you realize, "Next time, I should brake a little earlier." With a driving instructor, you secure a lesson every week, and then you make progress. If you do not secure that, you will never make it to your driving test. If you expect to get your driver's license by December 1st, you need to ensure that you regularly take driving lessons. It is the same with this.

Good point. Good comparison. I always like making comparisons when discussing a complex topic like AI. So indeed, now with driving lessons or the shift from being a mechanic to a controller, or looking at the expectations people had in 2000 and earlier when the internet was new. You do not know where it is headed, and the same applies here. You are relatively early in the learning curve, but you do not know where it will end.

No, and that means you cannot wait. The train will pass you by. And what role do you want to take? Well, you also have a diverse team with experience, so within Taurus, you can certainly make use of that in your training. And otherwise, of course, in other ways. Yes, I think it is smart. Look, if you consider the field of administration, in 2002, I had people employed who would just scan and transfer invoices. That was the job of an administrator. These were people aged 30 or 25 that I employed back then. Now, if you look at it, less than 20 years later, that entire function no longer exists. So, if you are talking about someone who's 50 now, it is all about computer technology. Nothing exciting. But really, if you are still scanning invoices now, you are way behind. It probably still happens at some companies. But normally, you receive it digitally, it is automatically analyzed, automatically linked to your ledger account. And then an automatic flow is set up for approval as well. You stamp it off with an app, and you are done, and the payment is set up. As an SME, you can do this with an e-bookkeeping package or tool, but that is already the new standard. So, you graduate, you have learned what needs to be done, but you only need to set it up once, and then you are already a controller. And the next step, as you easily say, is analyzing, "What is the standard?" And if you want to apply for a loan at the bank under 1 million, I think it is done in an Excel-like way. You upload your files, and for your mortgage, you submit some documents, but for a business, I assume it is the same. Because all those personal bank advisors are already gone. You do not even have them here. It is just an Excel sheet where your development is reviewed, your cash flow is checked. Based on that, an automatic calculation is made to see if you qualify for a loan. So, those roles are already gone, but that does not mean the people who did them do not have to transition to a new role. And as a company, you also have to go through that hoop together, thinking, "Darn, I am taking driving lessons," and you see someone looking, "Oh, you are taking driving lessons in a parking lot, that is funny. Oh, he just drove into a pole." Okay, well, that is funny, and you come back to the office, "Ah, that did not go well, did it?" "No, it did not go well, I hit a pole." "Okay," and you move on. Then someone else says, "Yeah, I am still fiddling with my horse and cart, but you are now driving comfortably on the road, that is nice. Can we learn that too?" But that process, our minds work slower because you have learned a trick that took many years to master, and many companies expect that by implementing it, you will suddenly turn into a superhuman. But your mind just works differently, you really need time. That is those 30 driving lessons, while the new tractor has already started, and you do not even know how to drive it, let alone have the experience to work with it. But do I need a big tractor or a smaller one for the job? You have done research, do you even know what suits your company best? Well, that is definitely something to consider in your assignment. How do you lay a solid foundation from the abstract to the present? Because otherwise, with this technology and the possibilities it offers, you will never get out of the starting blocks. That is the danger because it is so fast and all-encompassing.

Appendix I – Assessment FREEZE ChatGPT 4o

Assessment FREEZE ChatGPT – Assessor 1

Factual correctness of the LLM-generated texts

Qualitative

1. What is your impression of this analysis?

Below par. Factual incorrect items and missing pivotal points such as net congestion.

2. How relevant is the text that is written in this analysis?

Certain passages can be used to develop a good analysis but independently unusable.

3. What is/are the strongest parts in this text?

The broad market analysis. Especially in this case where the seller has little market input.

4. What is/are the weakest parts in this text?

The factual incorrectness.

5. How complete is the text?

Very limited.

6. What are important factors that need to be accounted for, when writing a text?

Touching the key items for the business case and that has not been done.

7. Is the text factual correct, compared to the original text that was used in the actual IM?

No.

Quantitative

1. Can you assess the clarity of this text on a scale from 1 to 10?

← Bad					Good →				
	x								

2. Can you assess the relevance of this text on a scale from 1 to 10?

← Bad					Good →				
	x								

3. Can you assess the structure of this text on a scale from 1 to 10?

← Bad					Good →				
				x					

4. What is your overall judgment of the text on a scale from 1 to 10?

← Bad					Good →				
		x							

Estimated time

1. How much time did you spend assessing this text? 15 minutes.
2. How much time would you expect to spend to make this text applicable for an IM? You can only use this text as an inspiration but not for an IM.

Appendix J – Assessment FREEZE Gemini

Assessment FREEZE Gemini – Assessor 1

Factual correctness of the LLM-generated texts

Qualitative

1. What is your impression of this analysis?

Below par. Factual incorrect items make the analysis not useful.

2. How relevant is the text that is written in this analysis?

The text is too commercially written for a company in this line of business. Tone of voice is not suitable.

3. What is/are the strongest parts in this text?

Offering a broader view.

4. What is/are the weakest parts in this text?

The factual incorrectness (Waalwijk instead of Harreveld, leading player) and the too commercial text.

5. How complete is the text?

Very limited.

6. What are important factors that need to be accounted for, when writing a text?

Factual correctness.

7. Is the text factual correct, compared to the original text that was used in the actual IM?

No.

Quantitative

1. Can you assess the clarity of this text on a scale from 1 to 10?

← Bad					Good →				
		X							

2. Can you assess the relevance of this text on a scale from 1 to 10?

← Bad					Good →				
		X							

3. Can you assess the structure of this text on a scale from 1 to 10?

← Bad	Good →

				x					

4. What is your overall judgment of the text on a scale from 1 to 10?

← Bad					Good →				
		x							

Estimated time

1. How much time did you spend assessing this text? 15 minutes.
2. How much time would you expect to spend to make this text applicable for an IM? You can only use this text as an inspiration but not for an IM.

Appendix K – Assessment DIAMOND ChatGPT 4o

Assessment DIAMOND ChatGPT – Assessor 1

Factual correctness of the LLM-generated texts

Qualitative

1. What is your impression of this analysis?

Below par. Crucial errors on the retail strategy.

2. How relevant is the text that is written in this analysis?

Many repetitions of less important facts.

3. What is/are the strongest parts in this text?

Limited strong parts.

4. What is/are the weakest parts in this text?

The factual accuracy and the number of repetitions of certain 'facts'.

5. How complete is the text?

Insufficient. It is a sequence of one-liners.

6. What are important factors that need to be accounted for, when writing a text?

Factual accuracy is most important and there are errors on this part.

7. Is the text factual correct, compared to the original text that was used in the actual IM?

No.

Quantitative

1. Can you assess the clarity of this text on a scale from 1 to 10?

← Bad					Good →				
		x							

2. Can you assess the relevance of this text on a scale from 1 to 10?

← Bad					Good →				
	x								

3. Can you assess the structure of this text on a scale from 1 to 10?

← Bad					Good →				
					x				

4. What is your overall judgment of the text on a scale from 1 to 10?

← Bad					Good →				
x									

Estimated time

1. How much time did you spend assessing this text? 20 minutes
2. How much time would you expect to spend to make this text applicable for an IM? Not usable for this purpose. So it needs to be complete rewritten.

Assessment DIAMOND ChatGPT – Assessor 2

Factual correctness of the LLM-generated texts

Qualitative

1. What is your impression of this analysis?
   My impression is that it provides a nice framework for organizing structure and information. However, the necessary relevant information is missing. It still provides a far from complete structure.
2. How relevant is the text that is written in this analysis?
   Most of the information is relevant. But it is too little for a comprehensive story.
3. What is/are the strongest parts in this text?
   The activities part. It goes right to the core of the activities of MBF. Hardly any irrelevant information is given.
4. What is/are the weakest parts in this text?
   The weakest part in the text is the branding analysis. In my opinion there is a lot to find about the branding, but the text out of Chat GPT has little substance on this piece.
5. How complete is the text?
   I am missing relevant information about product groups. Some of them are not mentioned at all.
6. What are important factors that need to be accounted for, when writing a text?
   Years and numbers are not always correct. It is especially important to check hard data.

7. Is the text factual correct, compared to the original text that was used in the actual IM?

   See point 6.

Quantitative

1. Can you assess the clarity of this text on a scale from 1 to 10?

← Bad					Good →				
					6				

2. Can you assess the relevance of this text on a scale from 1 to 10?

← Bad					Good →				
					6				

3. Can you assess the structure of this text on a scale from 1 to 10?

← Bad					Good →				
						7			

4. What is your overall judgment of the text on a scale from 1 to 10?

← Bad					Good →				
				5					

Estimated time

1. How much time did you spend assessing this text?

   4 hours

2. How much time would you expect to spend to make this text applicable for an IM?

   12 hours

Assessment DIAMOND ChatGPT – Assessor 3

Factual correctness of the LLM-generated texts

Qualitative

1. What is your impression of this analysis?

It starts off with information that is not true, but it tells it quite convincingly. The text itself is written in professional and clear language, the generated information is not all true or is exaggerated.

2. How relevant is the text that is written in this analysis?

Most parts are a bit too vague. Also, there is quite a lot of generated information that is not entirely true.

3. What is/are the strongest parts in this text?

USP's

Customers

4. What is/are the weakest parts in this text?

History and background. → lots of misinformation.

Market → short and too vague.

Suppliers → Some information is not correct.

5. How complete is the text?

It misses some important details, however, from only public sources the text is quite complete.

6. What are important factors that need to be accounted for, when writing a text?

Be concise. Be clear. Speak facts

7. Is the text factual correct, compared to the original text that was used in the actual IM?

To some extent. Some parts are factual correct. Other parts are not.

Quantitative

1. Can you assess the clarity of this text on a scale from 1 to 10?

← Bad					Good →				
								X	

2. Can you assess the relevance of this text on a scale from 1 to 10?

← Bad					Good →				
			X						

3. Can you assess the structure of this text on a scale from 1 to 10?

← Bad					Good →				
					X				

4. What is your overall judgment of the text on a scale from 1 to 10?

← Bad					Good →				
				X					

Estimated time

1. How much time did you spend assessing this text?

About 45 minutes

2. How much time would you expect to spend to make this text applicable for an IM?

At least a day to fact check, back up statements by data and rewrite some parts.

Appendix L – Assessment DIAMOND Gemini

Assessment DIAMOND Gemini – Assessor 1

Factual correctness of the LLM-generated texts

Qualitative

1. What is your impression of this analysis?

Very good.

2. How relevant is the text that is written in this analysis?

Very relevant.

3. What is/are the strongest parts in this text?

Factual accuracy and the structured presentation.

4. What is/are the weakest parts in this text?

The number of words (it has become a lengthy text).

5. How complete is the text?

Very complete, but this has to be fact-checked with the seller. Lots of new input that has to be verified.

6. What are important factors that need to be accounted for, when writing a text?

Factual accuracy is most important and there are errors on this part.

7. Is the text factual correct, compared to the original text that was used in the actual IM?

Cannot be compared due to new information (unverified), but appears to be accurate.

Quantitative

1. Can you assess the clarity of this text on a scale from 1 to 10?

← Bad					Good →				
								X	

2. Can you assess the relevance of this text on a scale from 1 to 10?

← Bad					Good →				

						X			

3. Can you assess the structure of this text on a scale from 1 to 10?

← Bad					Good →				
						x			

4. What is your overall judgment of the text on a scale from 1 to 10?

← Bad					Good →				
							x		

Estimated time

1. How much time did you spend assessing this text? 15 minutes
2. How much time would you expect to spend to make this text applicable for an IM? 3-4 hours (interview management and adapting the text accordingly).

Assessment DIAMOND Gemini – Assessor 2

Factual correctness of the LLM-generated texts

Qualitative

1. What is your impression of this analysis?
   Overall it is a good analysis. Relevant information is given and it is quite comprehensive. However, parts of the information is not right. For example, Gemini states that the company was founded by 2 persons. That is not the case. It was acquired by 2 persons who proceeded the company to what it is today.

2. How relevant is the text that is written in this analysis?

   Almost all the text is relevant. It is well-written and comprehensive.

3. What is/are the strongest parts in this text?

   The customer types. Gemini has provided a good characterization of the different customer types. This information is not available on a one-to-one basis in public information. Gemini has put together a good profile by connecting indirect information to customer types

4. What is/are the weakest parts in this text?

   Geographic customer distribution part. It is in my opinion too hollow. This part could be elaborated more in dept.

5. How complete is the text?

   It is very complete. There is in some parts more relevant information in it than in our own texts.

6. What are important factors that need to be accounted for, when writing a text?

   Specify the questions in depth to Gemini. Than the outcome will be more and more relevant.

7. Is the text factual correct, compared to the original text that was used in the actual IM?

   See point 1. But overall it is factual correct.

Quantitative

1. Can you assess the clarity of this text on a scale from 1 to 10?

← Bad									Good →
							8		

2. Can you assess the relevance of this text on a scale from 1 to 10?

← Bad									Good →
								9	

3. Can you assess the structure of this text on a scale from 1 to 10?

← Bad									Good →
							8		

4. What is your overall judgment of the text on a scale from 1 to 10?

← Bad					Good →				
							8		

Estimated time

1. How much time did you spend assessing this text?

   3 hours

2. How much time would you expect to spend to make this text applicable for an IM?

   12 hours

Assessment DIAMOND Gemini – Assessor 3

Factual correctness of the LLM-generated texts

Qualitative

1. What is your impression of this analysis?

Factually better than GPT. But DIAMOND is not a market leader. More written towards potential buyers. The text is written in a more commercial way where the GPT text is written more towards an analytical perspective.

2. How relevant is the text that is written in this analysis?

The text is relevant. However, I feel that some parts of the analysis are repeated too often.

3. What is/are the strongest parts in this text?

Introduction

Product

Market analysis

4. What is/are the weakest parts in this text?

Activities → too broad, not concise.

Structure

5. How complete is the text?

The text is quite complete. Some parts (sustainability, strategy)

6. What are important factors that need to be accounted for, when writing a text?

Factual, structure and also length of the text.

7. Is the text factual correct, compared to the original text that was used in the actual IM?

To some extent. Lot of emphasis on the business and marketing. Sometimes too extensive and structure wise not always logical.

Quantitative

1. Can you assess the clarity of this text on a scale from 1 to 10?

← Bad					Good →				
							x		

2. Can you assess the relevance of this text on a scale from 1 to 10?

← Bad					Good →				
							x		

3. Can you assess the structure of this text on a scale from 1 to 10?

← Bad					Good →				
		x							

4. What is your overall judgment of the text on a scale from 1 to 10?

← Bad					Good →				
					x				

Estimated time

1. How much time did you spend assessing this text?

45 minutes

2. How much time would you expect to spend to make this text applicable for an IM?

Around a day for fact checking and also removing the excess text.

Appendix M – Assessment SCRAP ChatGPT 4o

Assessment SCRAP ChatGPT – Assessor 4

Factual correctness of the LLM-generated texts

Qualitative

1. What is your impression of this analysis?
   - It is correct on certain items (less that Gemini), and has a flaw that it works to certain amounts of words per item. It therefore does not emphasize/elaborate where necessary
2. How relevant is the text that is written in this analysis?
   - It is relevant on activities and processes, it is not relevant on items which are not shown on the website.
3. What is/are the strongest parts in this text?
   - The summarization of certain texts are quite useful sometimes. I question how it found some customer/supplier data. In general this is not publicly available but still it does analyze some parts correct.
4. What is/are the weakest parts in this text?
   - The fact that is does provide words where it is not correct, therefore everything can be questioned.
5. How complete is the text?

- It is quite complete but incorrect, obviously it also lacks personal information/motives and a comparative setting.

6. What are important factors that need to be accounted for, when writing a text?
   - See answer GEMINI TTB
7. Is the text factual correct, compared to the original text that was used in the actual IM?
   - On subparts the text is factual correct, it is however surrounded by false analysis.

Quantitative

1. Can you assess the clarity of this text on a scale from 1 to 10?

← Bad									Good →
	x								

2. Can you assess the relevance of this text on a scale from 1 to 10?

← Bad									Good →
		x							

3. Can you assess the structure of this text on a scale from 1 to 10?

← Bad									Good →
x									

4. What is your overall judgment of the text on a scale from 1 to 10?

← Bad									Good →
		x							

Estimated time

1. How much time did you spend assessing this text?

   30 minutes
2. How much time would you expect to spend to make this text applicable for an IM?

   As much as you need to write a new IM

Assessment SCRAP ChatGPT – Assessor 1

Factual correctness of the LLM-generated texts

Qualitative

1. What is your impression of this analysis?
   Lesser than Gemini, but still the best so far.
2. How relevant is the text that is written in this analysis?
   All key points are being covered (for so far online available).
3. What is/are the strongest parts in this text?
   The combination of market intel and company information. It gives the impression of having a broad scope.
4. What is/are the weakest parts in this text?
   Sometimes overly simple texts and a bit too optimistic. Strong points of the company are being exaggerated.
5. How complete is the text?
   Not sufficiently to use as an IM.
6. What are important factors that need to be accounted for, when writing a text?
   To give an objective analysis (instead of a sales pitch)
7. Is the text factual correct, compared to the original text that was used in the actual IM?
   Much is correct, but the pitfalls are important (using the wrong name, calling the mooring in Almelo a port, or emphasizing that the facilities have a huge impact on regional economy or work possibilities).

Quantitative

1. Can you assess the clarity of this text on a scale from 1 to 10?

← Bad									Good →
								9	

2. Can you assess the relevance of this text on a scale from 1 to 10?

← Bad									Good →
						7			

3. Can you assess the structure of this text on a scale from 1 to 10?

← Bad									Good →
								9	

4. What is your overall judgment of the text on a scale from 1 to 10?

← Bad									Good →
		3							

Estimated time

1. How much time did you spend assessing this text? 34 minute
2. How much time would you expect to spend to make this text applicable for an IM? Days

Appendix N – Assessment SCRAP Gemini

Assessment SCRAP Gemini – Assessor 4

Factual correctness of the LLM-generated texts

Qualitative

1. What is your impression of this analysis?
   - On specific parts the analysis is very good, summarized but with enough detail, sometimes the analysis was way of with attention for details as if they were major developments. Although a lot of the basic information is included, Gemini sometimes explains this information wrongfully.
2. How relevant is the text that is written in this analysis?
   - Partially relevant, especially some summarized sentences could very well be used in an IM.
3. What is/are the strongest parts in this text?
   - Information about the activities of the company
4. What is/are the weakest parts in this text?
   - Explanations behind information such as locations and why it is used/important. Gemini comes up with an explanation what could read as a correct sentence (for some) but does not make the correct analysis.

5. How complete is the text?
   - Define complete, certain analysis are quite complete, sometimes this is off.
6. What are important factors that need to be accounted for, when writing a text?
   - These texts do not give a true comparative setting, and it obviously lacks the personal insights and motivations for a potential sales process.
7. Is the text factual correct, compared to the original text that was used in the actual IM?
   - On specific parts it is factual correct

Quantitative

1. Can you assess the clarity of this text on a scale from 1 to 10?

← Bad									Good →
			x						

2. Can you assess the relevance of this text on a scale from 1 to 10?

← Bad									Good →
				x					

3. Can you assess the structure of this text on a scale from 1 to 10?

← Bad									Good →
	x								

4. What is your overall judgment of the text on a scale from 1 to 10?

← Bad									Good →
			x						

Estimated time

1. How much time did you spend assessing this text?
   - 30 minutes
2. How much time would you expect to spend to make this text applicable for an IM?
   - Probably as long as you would work on building an IM yourself

Assessment SCRAP Gemini – Assessor 1

Factual correctness of the LLM-generated texts

Qualitative

1. What is your impression of this analysis?

   Even better than ChatGPT's version of SCRAP.

2. How relevant is the text that is written in this analysis?

   The differences between the two outlets is better explained than by ChatGPT.

3. What is/are the strongest parts in this text?

   The style of writing is business like. Not a simple sales pitch.

4. What is/are the weakest parts in this text?

   Information that is not online available is not in the text. These are for example the NDA-info.

5. How complete is the text?

   Not sufficient to use as an IM.

6. What are important factors that need to be accounted for, when writing a text?

   To give an objective analysis (instead of a sales pitch). This is a strongpoint.

7. Is the text factual correct, compared to the original text that was used in the actual IM?

   The available info is correct.

Quantitative

1. Can you assess the clarity of this text on a scale from 1 to 10?

← Bad									Good →
								9	

2. Can you assess the relevance of this text on a scale from 1 to 10?

← Bad									Good →
								9	

3. Can you assess the structure of this text on a scale from 1 to 10?

← Bad	Good →

								9	

4. What is your overall judgment of the text on a scale from 1 to 10?

← Bad									Good →
							8		

Estimated time

1. How much time did you spend assessing this text? 20 minutes
2. How much time would you expect to spend to make this text applicable for an IM? We would need an interview and non-disclosed info. For the rest this text is very usable.

Appendix O – Assessment WIRE ChatGPT 4o

Assessment WIRE ChatGPT – Assessor 5

Factual correctness of the LLM-generated texts

Qualitative

1. What is your impression of this analysis?
   Lots of content generated, 25% more words as Taurus text.
   Easy to read and also relevant. For example: USPs are described quite nicely and the strategy of the various specialized web shops and product configurators is also well presented.
2. How relevant is the text that is written in this analysis?
   The text is certainly relevant. But at the same time, it does read more like a story, rather than a fact-based professional text. Remarkably few facts, numbers and amounts. Example, at investment considerations:
   "Consistent Revenue Growth: (WIRE) has shown steady financial growth, demonstrating a resilient business model." Here you would normally expect some numbers or specific growth rates over a certain period of time. This does not enhance the credibility of the text and sometimes makes it one-liners.
3. What is/are the strongest parts in this text?

I am positively surprised by the fact that AI, for example, is also able to describe USPs. That is not something that can be found that way on a website it seems to me.

Going into previous, the USPs. A piece of text is generated that cannot be properly fact-checked. That does make it difficult to have a value judgment on correctness. USPs touts: wide product range, b2b and b2c customers, partnerships with manufacturers, online model, logistical edge and that with its online platform it is able to better understand the deeper customer needs.

4. What is/are the weakest parts in this text?

   Financial Performance is very limited, while quite a bit can be gleaned from public sources. There are no numbers in it either.

   Growth potential is very thin. Bears no relation to the growth plan we had made and is very important.

5. How complete is the text?

   The system always seems to give an answer and is not quick to point out that she is not sure. This is tricky, though.

6. What are important factors that need to be accounted for, when writing a text?

   The relevance and accuracy needs to be determined for every sentence.

7. Is the text factual correct, compared to the original text that was used in the actual IM?

   The big picture is pretty similar.

   In addition, it is more text, but less pointed and less fact based.

Quantitative

1. Can you assess the clarity of this text on a scale from 1 to 10?

← Bad									Good →
						X			

2. Can you assess the relevance of this text on a scale from 1 to 10?

← Bad									Good →
					x				

3. Can you assess the structure of this text on a scale from 1 to 10?

← Bad					Good →				
					x				

4. What is your overall judgment of the text on a scale from 1 to 10?

← Bad					Good →				
					x				

Estimated time

1. How much time did you spend assessing this text?

   1 hour

2. How much time would you expect to spend to make this text applicable for an IM?

   I would have to make it more fact based and therefore much more time is needed before I can finalize it to an IM. AI would help me and give me inspiration how to describe topics. But I still need to check the relevance, accuracy, and completeness. So more supportive than an alternative.

Assessment WIRE ChatGPT – Assessor 1

Factual correctness of the LLM-generated texts

Qualitative

1. What is your impression of this analysis?

   Good text. Yet on the commercial side.

2. How relevant is the text that is written in this analysis?

   Very relevant. You can only ask for more detail and more focus.

3. What is/are the strongest parts in this text?

   The complete image of the outer shell of the company.

4. What is/are the weakest parts in this text?

   The commercial and overly positive writing style and very limited internal information.

5. How complete is the text?

   Not sufficient to use as an IM. But a great step forward.

6. What are important factors that need to be accounted for, when writing a text?

To give an objective analysis (instead of a sales pitch). This is a weak point.

7. Is the text factual correct, compared to the original text that was used in the actual IM?

   The available info is correct.

Quantitative

1. Can you assess the clarity of this text on a scale from 1 to 10?

← Bad									Good →
							8		

2. Can you assess the relevance of this text on a scale from 1 to 10?

← Bad									Good →
							8		

3. Can you assess the structure of this text on a scale from 1 to 10?

← Bad									Good →
								9	

4. What is your overall judgment of the text on a scale from 1 to 10?

← Bad									Good →
							8		

Estimated time

1. How much time did you spend assessing this text? 30 minutes
2. How much time would you expect to spend to make this text applicable for an IM? We would need an interview and non-disclosed info. For the rest this text is very usable.

Appendix P – Assessment WIRE Gemini

Assessment WIRE Gemini – Assessor 5

Factual correctness of the LLM-generated texts

Qualitative

1. What is your impression of this analysis?

   In words 95% of our own IM, more succinctly so.

   Incredible that AI can determine the sale rationale - which it does say. AI cannot know that.

   Gemini more concrete, shorter bullets. More details. If it is correct then it increases credibility.
2. How relevant is the text that is written in this analysis?

   Pretty relevant. I see little text that does not belong to the case study. There is relatively much overlap through the topics though. But so is our own IM.
3. What is/are the strongest parts in this text?

   Good that SEA strategy is also written out. Incl Landing pages.

   Private label details - with explained that there are higher margins on this. Product categories are named, logically lacking the numbers to properly estimate the ratios.

   Proposition is pretty well described at business model.

   Good details on parts, for example: Their financial performance is impressive, with revenue growing from €46.6 million in 2020 to €132.7 million in 2022. Profit also saw significant increases, reaching €23 million in 2022. This consistent growth demonstrates their ability to capitalize on market opportunities and deliver strong returns.
4. What is/are the weakest parts in this text?

   Competitive playing field is very short, but it makes sense.

   Growth plan very thin. Also no market size determined and market shares missing.

   SWOT is missing. The risk or indirect exposure of Solar is yet lifted out very clear.
5. How complete is the text?

   To get an overall picture it is pretty impressive. To convince less because you are less guided than by an IM. And also important facts are missing like for example the how energy efficient the warehouse in Enschede is. The construction of the new warehouse in Almelo is missing. And also details that AI cannot know such as internal KPIs and relationship (statistics of SEA, orders, customer satisfaction). No analyses of number of customers, concentration, turnover per order, etc. Similar details are missing on the procurement side.
6. What are important factors that need to be accounted for, when writing a text?

Very broad question. But when is comes to AI: The relevance and accuracy needs to be determined for every sentence.

7. Is the text factual correct, compared to the original text that was used in the actual IM?

   Yes but not factually complete. It is written more generically and narratively than an IM, so it is not easily wrong at its core - but certainly less relevant.

Quantitative

1. Can you assess the clarity of this text on a scale from 1 to 10?

← Bad					Good →				
						x			

2. Can you assess the relevance of this text on a scale from 1 to 10?

← Bad					Good →				
					x				

3. Can you assess the structure of this text on a scale from 1 to 10?

← Bad					Good →				
						x			

4. What is your overall judgment of the text on a scale from 1 to 10?

← Bad					Good →				
					x				

Estimated time

1. How much time did you spend assessing this text?

   1 hour

2. How much time would you expect to spend to make this text applicable for an IM?

   Cannot answer. More general parts can be quite easily integrated. But text need to be enriched with data from internal client sources.

Assessment WIRE Gemini – Assessor 1

Factual correctness of the LLM-generated texts

Qualitative

1. What is your impression of this analysis?

   Very good text (as far as information is publicly available).
2. How relevant is the text that is written in this analysis?

   Very relevant. You can only ask for more detail and more focus.
3. What is/are the strongest parts in this text?

   The complete image of the outer shell of the company.
4. What is/are the weakest parts in this text?

   Very limited internal information. The key issue is the composition of electrical circuit boards. But the composition part is not mentioned.
5. How complete is the text?

   Not sufficient to use as an IM. But a great step forward.
6. What are important factors that need to be accounted for, when writing a text?

   To give an objective analysis (instead of a sales pitch). This is a strongpoint.
7. Is the text factual correct, compared to the original text that was used in the actual IM?

   The available info is correct.

Quantitative

1. Can you assess the clarity of this text on a scale from 1 to 10?

← Bad									Good →
								9	

2. Can you assess the relevance of this text on a scale from 1 to 10?

← Bad									Good →
								9	

3. Can you assess the structure of this text on a scale from 1 to 10?

← Bad	Good →

								9	

4. What is your overall judgment of the text on a scale from 1 to 10?

← Bad									Good →
							8		

Estimated time

1. How much time did you spend assessing this text? 30 minutes
2. How much time would you expect to spend to make this text applicable for an IM? We would need an interview and non-disclosed info. For the rest this text is very usable.

Appendix Q – Assessment OMNIA ChatGPT 4o

Assessment OMNIA ChatGPT – Assessor 2

Factual correctness of the LLM-generated texts

Qualitative

1. What is your impression of this analysis?

Overall it is a good analysis. Relevant information is given and it is quite comprehensive. However some relevant information (for example the different branches) is not completely correct. .

2. How relevant is the text that is written in this analysis?

All the information is relevant. Some parts could be elaborated more.

3. What is/are the strongest parts in this text?

   Activities

4. What is/are the weakest parts in this text?

   Strategy; this part is very briefly.

5. How complete is the text?

   It is quite complete.

6. What are important factors that need to be accounted for, when writing a text?

   Specify the questions in dept to Chat gpt. Than the outcome will be more and more relevant.

7. Is the text factual correct, compared to the original text that was used in the actual IM?

   It is overall factual correct. On the branches part it is not completely correct. Some parts are missing.

Quantitative

1. Can you assess the clarity of this text on a scale from 1 to 10?

← Bad									Good →
					7				

2. Can you assess the relevance of this text on a scale from 1 to 10?

← Bad									Good →
						8			

3. Can you assess the structure of this text on a scale from 1 to 10?

← Bad									Good →
					7				

4. What is your overall judgment of the text on a scale from 1 to 10?

← Bad									Good →
						8			

Estimated time

1. How much time did you spend assessing this text?

   1,5 hours

2. How much time would you expect to spend to make this text applicable for an IM?

   30 hrs

Assessment OMNIA ChatGPT – Assessor 6

Factual correctness of the LLM-generated texts

Qualitative

1. What is your impression of this analysis?
   Actually pretty amazing how GPT has made it possible to present a text which – in broad terms – describes the business. I am impressed.
2. How relevant is the text that is written in this analysis?
   parts of the text are surely relevant, other parts of the text are more descriptive in nature than our IM. For example; GPT speaks about technology being a part of the strategy of the company. Yes, as far as this is needed to operate efficiently, but the company is far from a pioneer in this field.
3. What is/are the strongest parts in this text?
   Especially the parts about the general market and external market analysis are in my opinion the stronger parts of the text. Maybe because this is not company-specific?
4. What is/are the weakest parts in this text?
   GPT uses words and phrases like "(Company name)….is an industry leader, and…. Delivering exceptional results". This is for both not the case. Besides that this is not true, I would rather not use this wording anyway.
5. How complete is the text?
   Considering the lack of input, except prompts, the text is pretty complete. When knowing the business one can see the flaws, but in more generic terms it is pretty complete.
6. What are important factors that need to be accounted for, when writing a text?
   What GPT – probably understandable – cannot account for is the real context of the business. The background information and emotion. GPT seems not able, at this point, to make a text truly personal. To me it seems this text could be used for more secondment business like the company. Not a tailormade text.
7. Is the text factual correct, compared to the original text that was used in the actual IM? Not all parts are factual correct. GPT refers to an add-on whereby the company increased its capacity with 1.500 employees whereas in the IM it says 300 employees.

GPT positions the company as a company with nationwide coverage, whereas this is an overstatement.

Quantitative

1. Can you assess the clarity of this text on a scale from 1 to 10?

← Bad									Good →
						7			

2. Can you assess the relevance of this text on a scale from 1 to 10?

← Bad									Good →
					6				

3. Can you assess the structure of this text on a scale from 1 to 10?

← Bad									Good →
						7			

4. What is your overall judgment of the text on a scale from 1 to 10?

← Bad									Good →
					6				

Estimated time

1. How much time did you spend assessing this text?
   overall I think about 30 minutes
2. How much time would you expect to spend to make this text applicable for an IM?
   The generic parts can be copied into the IM. The company-specific parts need to be questioned and verified by the company. This will most likely take up several hours (Q&A sessions) to make all this factual right.

Appendix R – Assessment OMNIA Gemini

Assessment OMNIA Gemini – Assessor 2

Factual correctness of the LLM-generated texts

Qualitative

1. What is your impression of this analysis?

   Overall it is a good analysis. Relevant information is given and it is quite comprehensive. However, some relevant information (for example about the different branches of OMNIA) is not given.
2. How relevant is the text that is written in this analysis?

   All the information is relevant. Some parts could be elaborated more.
3. What is/are the strongest parts in this text?

   Distinctive strengths.
4. What is/are the weakest parts in this text?

   Strategy; this part is very briefly.
5. How complete is the text?

   It is quite complete.
6. What are important factors that need to be accounted for, when writing a text?

   Specify the questions in dept to Gemini. Then the outcome will be more and more relevant.
7. Is the text factual correct, compared to the original text that was used in the actual IM?

   It is factual correct. Some parts are missing.

Quantitative

1. Can you assess the clarity of this text on a scale from 1 to 10?

← Bad									Good →
						8			

2. Can you assess the relevance of this text on a scale from 1 to 10?

← Bad									Good →
						8			

3. Can you assess the structure of this text on a scale from 1 to 10?

← Bad									Good →
					7				

4. What is your overall judgment of the text on a scale from 1 to 10?

← Bad									Good →
						8			

Estimated time

1. How much time did you spend assessing this text?

   1,5 hours

2. How much time would you expect to spend to make this text applicable for an IM?

   30 hours

Assessment OMNIA Gemini – Assessor 6

Factual correctness of the LLM-generated texts

Qualitative

1. What is your impression of this analysis?

   Compared to GPT the text by Gemini seems more tailored at the actual business case of the company. It is speaking less in broad terms and seems to generate more company-specific details.

2. How relevant is the text that is written in this analysis?

   The text seems relevant for the analysis. Some aspects are less applicable. Gemini refers to a market trend in cybersecurity and data analytics. Although relevant, less for the company.

3. What is/are the strongest parts in this text?

   The part about activities and history is well written considering an AI tool is generating this

4. What is/are the weakest parts in this text?

   Some parts are not mentioned in the IM. Gemini refers to equality as a core value of the company. Of course, the company promotes equality, but nowhere in the IM is this mentioned. To me, this mention was a surprise.
5. How complete is the text?

   Fairly complete
6. What are important factors that need to be accounted for, when writing a text?

   The context of the business and understanding the competitive distinct capabilities of a business. Knowing these enables you to understand the financials better.
7. Is the text factual correct, compared to the original text that was used in the actual IM?

   For the most part, I think it is correct, some aspects are not mentioned in the IM but are generated by Gemini. Especially these parts are the parts where I seem to have the most doubts.

Quantitative

1. Can you assess the clarity of this text on a scale from 1 to 10?

← Bad									Good →
							8		

2. Can you assess the relevance of this text on a scale from 1 to 10?

← Bad									Good →
						7			

3. Can you assess the structure of this text on a scale from 1 to 10?

← Bad									Good →
						7			

4. What is your overall judgment of the text on a scale from 1 to 10?

← Bad									Good →
						7			

Estimated time

1. How much time did you spend assessing this text?

   About half an hour

2. How much time would you expect to spend to make this text applicable for an IM?

   See comment GPT assessment