



# Prompt Engineering and Management Decision Making: Optimizing the Nexus of Human and Artificial Intelligence

Timo Sirch

Dissertation written under the supervision of professor Peter  
Rajsingh

Dissertation submitted in partial fulfilment of requirements for the MSc in  
Management with a specialization in Strategy, Entrepreneurship &  
Impact, at the Universidade Católica Portuguesa, 05.01.2026.

# Abstract

Managers increasingly use large language models to support early phases of decision making, yet little empirical evidence exists on how prompt design influences the quality of AI-supported problem structuring. This study examined how different prompt engineering strategies affect the quality of problem-structuring outputs generated by a large language model. Using a controlled mixed-design experiment, three prompting strategies (baseline, instruction, and decomposition prompting) were applied to standardized managerial cases representing different levels of problem structure. The resulting outputs were evaluated by a panel of practitioners using rating-based and ranking-based assessments across multiple quality dimensions, complemented by qualitative analysis.

The analysis finds systematic differences across prompting strategies and problem types. Baseline prompting consistently results in the worst problem structure quality, with decomposition prompting closely outperforming instruction prompting. While this effect increases as the problems become more complex and ambiguous, relative effectiveness measured through ranking of prompting strategies, remains stable across cases.

Overall, the study demonstrates that prompt engineering functions as a managerial design choice that meaningfully shapes AI-supported problem structuring. By improving the explicitness and organization of analytical reasoning, structured prompts can enhance the usefulness of AI outputs at a critical early stage of managerial decision making.

**Keywords:** Prompt engineering, large language models, AI-supported problem structuring, managerial decision making, strategic problem structuring, human–AI collaboration, experimental design, evaluation of AI outputs, decision quality

**Title:** Prompt Engineering and Management Decision Making: Optimizing the Nexus of Human and Artificial Intelligence

**Author:** Timo Sirch

# Sumário

Os gestores utilizam mais modelos linguísticos de grande dimensão para apoiar as fases iniciais da tomada de decisões, mas existem poucas evidências empíricas sobre como o design de prompts influencia a qualidade da estruturação de problemas apoiada por IA. Este estudo examinou como diferentes estratégias de engenharia de prompts afetam a qualidade dos resultados de estruturação de problemas gerados por um modelo linguístico de grande dimensão. Utilizando um experimento controlado de design misto, foram aplicadas três estratégias de prompt (base, instrução e decomposição) a casos de gestão padronizados que representam diferentes níveis de estrutura do problema. Os resultados foram avaliados por um painel de profissionais através de avaliações por classificação e por ordenação, complementadas por análise qualitativa.

A análise revela diferenças sistemáticas entre as estratégias de prompt e os tipos de problemas. O prompt de base resulta consistentemente na menor qualidade de estruturação de problemas, enquanto o prompt de decomposição apresenta o melhor desempenho, seguido do prompt de instrução. Embora este efeito se intensifique à medida que os problemas se tornam mais complexos e ambíguos, a eficácia relativa das estratégias de prompt, medida através da ordenação, permanece estável entre os casos.

De modo geral, o estudo demonstra que a engenharia de prompts funciona como uma escolha de design gerencial que molda significativamente a estruturação de problemas apoiada por IA. Ao melhorar a explicitação e a organização do raciocínio analítico, os prompts estruturados podem aumentar a utilidade dos resultados da IA num estágio inicial crítico da tomada de decisões gerenciais.

**Palavras-chave:** Engenharia de prompts, grandes modelos de linguagem, estruturação de problemas apoiada por IA, tomada de decisões gerenciais, estruturação estratégica de problemas, colaboração humano-IA, projeto experimental, avaliação dos resultados da IA, qualidade das decisões

**Título:** Engenharia de prompts e tomada de decisões gerenciais: otimizando a conexão entre inteligência humana e artificial

**Autor:** Timo Sirch

## Table of Contents

|  |    |
|--|----|
| 1. Introduction.....   | 7  |
| 2. Literature Review.....                                      | 8  |
| 2.1 Introduction .....   | 8  |
| 2.2 From Artificial Intelligence to Large Language Models..... | 8  |
| 2.2.1 Machine Learning .....                                   | 8  |
| 2.2.2 Deep Learning.....                                       | 9  |
| 2.2.3 Generative Artificial Intelligence.....                  | 9  |
| 2.2.4 Large Language Models (LLMs).....                        | 9  |
| 2.2.5 Managerial relevance of large language models.....       | 10 |
| 2.3 Human–AI Collaboration.....                                | 10 |
| 2.3.1 Concept and Principles .....                             | 10 |
| 2.3.2 Benefits and Opportunities .....                         | 11 |
| Productivity and task performance .....                        | 11 |
| Collaboration and communication.....                           | 11 |
| Long-term organizational outcomes .....                        | 11 |
| 2.3.3 Risks and Challenges .....                               | 12 |
| Performance Risks .....  | 12 |
| Cognitive Risks.....   | 12 |
| Organizational and Ethical Risks.....                          | 12 |
| 2.3.4 Managerial implications of human–AI collaboration.....   | 13 |
| 2.4 The Role of Prompt Engineering .....                       | 13 |
| 2.4.1 Concept and Foundations.....                             | 13 |
| 2.4.2 Prompting Strategies.....                                | 14 |
| Baseline-prompting.....  | 14 |
| Instruction Prompting .....                                    | 14 |
| Decomposition Prompting .....                                  | 15 |
| 2.4.3 Evaluating and Interpreting Prompt Outcomes.....         | 15 |
| 2.4.4 Managerial Capabilities of Prompt Engineering.....       | 15 |
| 2.5 Strategic Problem Structuring .....                        | 16 |
| 2.5.1 Why Managerial Problems Fail .....                       | 16 |
| 2.5.2 Types of managerial problems .....                       | 17 |

|  |    |
|--|----|
| Well-structured problems .....                                   | 17 |
| Ill-structured problems.....                                     | 17 |
| Wicked-structured problems .....                                 | 17 |
| 2.5.3 Effective Problem Structuring .....                        | 18 |
| 2.5.4 Evaluation Dimensions .....                                | 18 |
| Logical Coherence .....  | 18 |
| Actionability .....  | 18 |
| Comprehensiveness (MECE).....                                    | 18 |
| Depth of insight.....  | 19 |
| 2.5.5 Managerial Value of AI-Augmented Problem Structuring ..... | 19 |
| 2.6 Research Gap.....  | 19 |
| 2.6.1 Purpose and scope of the literature review .....           | 19 |
| 2.6.2 Relevance and study positioning .....                      | 20 |
| 3. Methodology .....   | 21 |
| 3.1 Research Design.....   | 21 |
| 3.2 Experimental Setup .....                                     | 21 |
| 3.3 Prompt Design.....   | 22 |
| 3.4 Operationalization of Prompting Strategies .....             | 23 |
| Baseline-prompting.....  | 23 |
| Instruction Prompting .....                                      | 23 |
| Decomposition Prompting .....                                    | 23 |
| 3.5 Prompt–Problem Matrix.....                                   | 24 |
| 3.6 Evaluation Framework .....                                   | 24 |
| 3.7 Data Analysis.....   | 25 |
| 3.8 Validity and Reliability.....                                | 26 |
| 4. Quantitative Analysis .....                                   | 27 |
| 4.1 Descriptive statistics.....                                  | 27 |
| 4.2 Model specification and estimation approach.....             | 28 |
| 4.3 Regression Results .....                                     | 29 |
| Effects of prompting strategies and case heterogeneity.....      | 30 |
| Evaluation dimensions .....                                      | 30 |
| 4.4 Ranking Analysis .....                                       | 30 |
| Descriptive ranking results .....                                | 31 |

|  |    |
|--|----|
| Interpretation and consistency with rating results ..... | 32 |
| 5. Qualitative Analysis .....                            | 33 |
| 5.1 Logical Coherence.....                               | 33 |
| 5.2 Comprehensiveness (MECE) .....                       | 34 |
| 5.3 Actionability .....                                  | 35 |
| 5.4 Depth of Insight.....                                | 36 |
| 6. Discussion .....                                      | 38 |
| 6.1 Why Structured Prompting Works .....                 | 38 |
| 6.2 Practical Implications and Boundary Conditions.....  | 39 |
| 7. Conclusion .....                                      | 39 |
| 8. Limitations and Future Research .....                 | 40 |
| 9. Bibliography .....                                    | 42 |
| 10. Appendix .....                                       | 47 |
| Test-Cases.....  | 47 |
| Prompt Inputs and Model Outputs .....                    | 53 |
| Baseline.....  | 53 |
| P01 Input:.....  | 53 |
| O01 Output.....  | 54 |
| P04 Input.....   | 55 |
| O04 Output.....  | 56 |
| P07 Input.....   | 57 |
| O07 Output.....  | 58 |
| Instruction Prompting .....                              | 59 |
| P02 Input.....   | 59 |
| O02 Output.....  | 60 |
| P05 Input.....   | 61 |
| O05 Output.....  | 63 |
| P08 Input.....   | 64 |
| O08 Output.....  | 65 |
| Decomposition Prompting .....                            | 66 |
| P03 Input.....   | 66 |
| O03 Output.....  | 68 |
| P06 Input.....   | 70 |

|                 |    |
|-----------------|----|
| O06 Output..... | 71 |
| P09 Input.....  | 73 |
| O09 Output..... | 75 |

# 1. Introduction

AI is a major transformative force in contemporary life. Its impacts are being felt in media, education, healthcare, and, most notably, the professional and corporate worlds. For most of recent history it was thought that low paying jobs were especially vulnerable to being replaced by AI, but it turns out that high-paying, skilled occupations are just as threatened (Eloundou, Manning, Mishkin, & Rock, 2023). In order to survive in this new landscape, people will need to adapt and use AI to not be left behind.

In point of fact, the real value driver of this technology lies in human-machine collaboration. This relates not only to outsourcing certain tasks but also to how people and AI work together to improve productivity and decision making. (Dell'Acqua, Fuller, Kim, & Kleinbaum, 2023). As a result, it's becoming increasingly important not necessarily to do all tasks yourself, but rather to optimize how and for what your AI agent can help you.

This has led to a new emerging field: Prompt Engineering. This industry is one of the fastest growing sectors in North America (according to market.us estimated by to grow 33.9% between 2025-2034) and highlights the importance of the process of designing and optimizing prompts to obtain relevant and accurate responses from LLMs (Fortune Business Insights, 2024).

That high routine and repetitive tasks are easily outsourced to AIs is evident. It becomes more interesting when we look at non-routine tasks with high cognitive complexity. Here we also find the highest resistance to outsourcing and encounter key issues such as trust and accuracy (Tinguely, Lee, & He, 2023). Optimally, guiding AIs helps to solve complex business decisions by clearly communicating a relevant and logical coherent structure, which gives actionable and relevant solutions to a specific problem. This leads to the research question of this dissertation which examines how different prompt engineering strategies influence the quality of AI generated problem structures in realistic business cases.

## 2. Literature Review

### 2.1 Introduction

This literature review grounds the thesis in a conceptual foundation by situating large language models within the broader landscape of artificial intelligence. It highlights the importance of managers mastering human–AI collaboration and examines prompt engineering as the key mechanism for guiding large language models. Building on this, the review focuses on problem structuring as a central lever for managerial support and outlines key factors of effective problem structuring while discussing how large language models can contribute to this process.

### 2.2 From Artificial Intelligence to Large Language Models

Artificial intelligence refers to computer systems or software that are designed to perform tasks which previously required human intelligence (Coursera, 2024). From a management perspective, AI is primarily used to augment human capabilities like automating repetitive or data-intensive tasks in order to improve decision making and to generate value (Quiroz-Vázquez & Goodwin, n.d.). As an umbrella term, artificial intelligence includes several subfields, most importantly machine learning and deep learning.

#### 2.2.1 Machine Learning

Machine learning algorithms have the ability to detect patterns in large data sets quickly and learn to make predictions by processing data rather than explicit programming instructions (McKinsey & Company, 2024a). While machine learning includes a broad range of data driven algorithms, some approaches rely on particularly complex model structures, leading to deep learning as a more specialized class of machine learning methods.

### 2.2.2 Deep Learning

Deep Learning uses neural networks, similar to neurons in the human brain, to capture statistical relationships in varying contexts, process this input through many iterations, and finally develop an understanding of increasing levels of complexity. This leads the AI to make iterative improvements on the level of sophistication of its predictions (McKinsey & Company, 2024a).

### 2.2.3 Generative Artificial Intelligence

Generative Artificial Intelligence can be seen as a class of deep learning models that use extremely large neural networks called large language models. While generative approaches also exist for other data like images or videos, large language models primarily focus on generating and interpreting text (McKinsey & Company, 2024b).

### 2.2.4 Large Language Models (LLMs)

Large Language Models represent a subcategory of generative AI and are typically composed of deep neural networks comprising billions of parameters, enabling them to generate, summarize and reason with language. Core mechanisms underlying LLMs include pre-training, during which the model is trained on vast amounts of unlabeled text to learn general linguistic patterns, fine-tuning which adapts the LLM for specific tasks and in-context-learning where models adjust their behavior based solely on the prompt without updating model parameters (Han et al., 2024).

The literature commonly distinguishes four main categories of LLMs. Text-Based or narrow LLMs are trained and operated exclusively on textual data (Stryker, n.d.) while instruction-based are able to follow human instructions or prompts more reliably (Bergmann, n.d.). Multimodal LLMs go beyond just text and are capable of processing text, images and audio or a combination of the three. Domain-specific large language models are trained to focus on a specific field and have a narrower application scope, usually for a particular industry like law or finance (Varughese, n.d.).

## 2.2.5 Managerial relevance of large language models

For managers, the proper use of large language models becomes increasingly important as a strategic tool. It changes the landscape of the managerial workflow, shifting emphasis away from implementing and executing to orchestration, focusing on problem framing, interpretation and oversight, rather than manual analysis (McKinsey & Company, 2024c).

## 2.3 Human–AI Collaboration

Building on their growing managerial relevance, the effectiveness of large language models ultimately depends on the nature of human–AI interaction and collaboration in decision making contexts.

### 2.3.1 Concept and Principles

Human–AI collaboration is grounded in core principles that define how humans and intelligent systems work together and complement each other in the context of task execution and decision making. Individual and collective intelligence refer to a task being done entirely by one or multiple humans without support from intelligent technologies. Automated intelligence describes tasks or processes that are fully executed by intelligent technologies without any input or guidance from a human in the decision making. Augmented intelligence or augmented collective intelligence is the situation in which one or multiple humans work collaboratively together with intelligent technologies to execute tasks, make decisions or accelerate cognitive and analytical work (Kolbjørnsrud, 2024). With the growing development of artificial intelligence and large language, it becomes increasingly important to learn how to design collaborative, augmented workflows that enable adaptive interaction between humans and AI-Systems (Berretta et al., 2023).

### 2.3.2 Benefits and Opportunities

The quick advancements of artificial intelligence have had a significant impact on managerial work.

#### *Productivity and task performance*

By gaining access to information, insights and alternative perspectives quickly, LLMs help leaders to make better-informed, data-driven and higher quality decisions (Varughese, n.d.). They can improve managerial productivity by automating many time-intensive and difficult tasks such as analyzing and summarizing large amounts of data or reports, which frees up time for managers to focus on more strategic tasks (Dell'Acqua et al., 2023). Importantly, generative AI shows an increase in performance across many different tasks, not only for simple or routine tasks.

#### *Collaboration and communication*

Empirical evidence shows that humans working in collaboration with AI can significantly outperform humans working independently, especially in creative tasks. Even when humans outperform artificial intelligence in isolation, the evidence suggests that working together with AI further improves the output (Vaccaro, Almaatouq, & Malone, 2024). LLMs also improve how managers communicate. The generated outputs are usually more positive in language and are characterized by simpler language compared to human outputs. Standardized, clear, and consistent communication is very important in organizations (Przegalinska et al., 2025).

#### *Long-term organizational outcomes*

The benefits of AI integration in companies are also evident in their long-term success. Firms that combine organizational learning with AI specific education are up to 80% more effective in navigating external uncertainty like technological or regulatory change and shifting market conditions. They also deal better with knowledge loss due to employee turnover compared to organizations with limited AI education. Financially, companies with a better understanding of

AI are 1.4 times more likely to report financial gains, further showing the benefits of AI in the workplace (Boston Consulting Group, 2024).

### 2.3.3 Risks and Challenges

Despite the potential of human–AI collaboration for managers and organizations, its implementation is accompanied by a range of risks and challenges that must be considered.

#### *Performance Risks*

Evidence suggests that while collaboration can outperform either human or AI-Systems in isolation, the best standalone performer for either performance on average is significantly better than human-AI combination. In tasks that require a lot of decision-making performance losses have also been observed (Vaccaro, Almaatouq, & Malone, 2024).

#### *Cognitive Risks*

Automation bias, where humans overly trust artificial intelligence outputs, can also significantly harm the joined output from humans and AI and reduce critical reflection (Romeo et al., 2025). People may rely too much on AI-Systems, accepting AI suggestions at face value without further analysis. Conversely, humans can also trust AI too little, ignoring input just because of an adverse attitude towards automation. Misalignment between humans and AI decreases overall performance substantially, especially when monitoring or coordination fails (Vaccaro, Almaatouq, & Malone, 2024).

#### *Organizational and Ethical Risks*

Organizations also face ethical risks as a result of unclear accountability, lack of transparency, and weak oversight in AI-driven decisions. Employees might shift blame to the AI outputs or even completely outsource ethical judgement to this technology. This can be dangerous when the ethics of the developers don't align with the users that train the system. AI adoption can also trigger anxiety related to job loss due to perceived risks of replacement (Heyder, 2023).

These effects may be amplified when culture and structure are in conflict within organizations increasing bias and eroding trust if AI-Systems are integrated without proper ethical safeguards (Bankins, 2024).

Taken together, these risks underline that realizing the benefits of human–AI collaboration requires deliberate managerial choices regarding task design, oversight, and integration.

### 2.3.4 Managerial implications of human–AI collaboration

The effective use of artificial intelligence by managers is crucial to realizing the demonstrated benefits of AI. This requires a solid understanding of how large language models function and how they can be applied in practice. To maximize the value of AI systems, it is important to understand and anticipate system behavior, establish trust, control outputs, and make sound decisions based on the generated information. These constitute four core conditions for managers to capture value from AI-Systems (Berretta et al., 2023). Building a strong practical competence in interacting with LLMs enables managers to integrate this technology most effectively into their workflow (Boston Consulting Group, 2024).

## 2.4 The Role of Prompt Engineering

Managers have the ability to interact with and guide large language models effectively through prompt engineering. This enables them to structure problems, guide analysis, and translate managerial intent into actionable outputs.

### 2.4.1 Concept and Foundations

Conceptually, prompt engineering is defined as the process of producing and optimizing inputs that guide large language models to have more accurate, relevant, and controllable outputs (Pandit, 2024). It acts as a practical interface between human intent and AI reasoning,

translating goals, tone, and context into structured language that models can interpret effectively (Google Cloud, n.d.).

## 2.4.2 Prompting Strategies

Because the same task can be expressed to a large language model in multiple ways, different strategies have emerged to guide the model to the desired outcome. Different prompt formulations can vary in terms of how explicitly they define objectives, structure a task, provide context or constrain the reasoning process of the model. While the literature describes a broad landscape of different prompt strategies, this thesis focuses on three main strategies (Sahoo et al., 2024).

### *Baseline-prompting*

Baseline-prompting represents an unstructured mode of interaction with large language models. This requires the AI to rely primarily on its general pre-trained capabilities to interpret and respond to tasks. It decides how to frame the problem, structure its reasoning, and organize its output without any human guidance. As such, it provides insight into how LLMs approach managerial problem structuring (Bonna et al., 2025).

### *Instruction Prompting*

Instruction Prompting refers to a strategy where explicit task instructions are provided to guide the model's response. This directs the large language model toward a specific objective, output format, or analytical focus, rather than relying solely on its general pre-trained behavior. It can be thought of as an umbrella category for any form of task guidance, including role specification, output constraints, or stylistic instructions. As a result, it reduces ambiguity in the interpretation of tasks and aligns the output closer with the managerial intent (Schulhoff, 2025).

### *Decomposition Prompting*

Decomposition Prompting breaks a complex task down into smaller, sequential, or logically distinct sub-tasks. Rather than only defining the overall objective, this prompt strategy guides the large language models by specifying intermediate steps or analytical components. This moves part of the problem-structuring effort from the model to the prompt, making the reasoning process more transparent. By reducing cognitive complexity, the AI-System can narrow its focus to each step and support more systematic analysis. Compared to instruction prompting, decomposition prompting goes a step further by structuring the reasoning process itself, not just the expected output or task goal (Kuka, 2024).

### 2.4.3 Evaluating and Interpreting Prompt Outcomes

Prompt design has a significant influence on the output of large language models. Empirical evidence suggests that changes in the structural design of prompts can introduced severe bias. Variables like phrasing, response order or labeling can systematically influence model outputs (Brucks & Toubia, 2025). Even small changes in prompt design can significantly alter LLM answer choices and go so far as to alter the internal decision-making processes of LLMs, potentially leading to output inaccuracies (Sun & Kok, 2025). However, Neutralizing or masking specific words or connotations can in turn improve performance consistency (Chadimová, Jurášek, & Kliegr, 2024). These findings highlight that prompt design can influence model outputs significantly. Therefore, the effective use of large language models ultimately depends on managers knowing how to not only interpret results but also being proficient in designing prompts and choose suitable prompting strategies.

### 2.4.4 Managerial Capabilities of Prompt Engineering

Sophisticated knowledge of prompt engineering is becoming an important skill for managers. It enables them to better benefit from the capabilities of AI-Systems and get the most useful answer for their problem, without the need for much IT or coding knowledge. They can just use natural language to guide the LLM which makes it far more accessible for managers to learn (McKinsey & Company, 2024d).

Importantly, from a resource-based perspective technologies alone do not create a competitive advantage, because they can be easily adopted by other firms (Barney et al., 2021). However, if used complementarily with managers, who are able to apply this technology in decision-making contexts, a real competitive advantage can arise.

A well-designed prompt improves accuracy by reducing misinterpretations and hallucinations, increases consistency across quality, structure, relevance, and tone. In this context, prompt engineering can be thought of as a managerial capability that enables managers to translate ill-structured problems into effective AI interactions and to systematically extract decision-relevant insights from otherwise generic AI tools. Clear prompting can also save time and cost by reducing repeated clarification and rework with the LLM (Anthropic, 2024).

## 2.5 Strategic Problem Structuring

While prompt engineering without a doubt is an important skill for managers to have, it is a means to an end. The goal of a prompt is the quality and usefulness of the output of the large language model. In a managerial context, the primary objective is rarely to generate final decisions or recommendations, but to support the manager in the decision-making process by initially structuring complex problems. Therefore, a key value proposition of large language models in the managerial context is the extent to which different prompt strategies enable LLMs to produce a high-quality strategic problems structure.

### 2.5.1 Why Managerial Problems Fail

Diving straight into solving a problem is rarely the right approach. Clarifying objectives, constraints and key drivers before the solution evaluation begins is a critical step in the beginning of a project and usually the responsibility of managers. Helping leaders clarifying the underlying problem-drivers before beginning to fix the problem can make the entire process more efficient and save time. Studies show that many projects fail because people don't take the time to examine problems from all angles and explore their complexities (Binder & Watkins, 2024).

## 2.5.2 Types of managerial problems

This shows that, in most managerial problems, the main difficulty lies not in developing the solution but in structuring the problem, depending on the problem type.

### *Well-structured problems*

A well-structured problem is characterized by having a clearly defined goal, known constraints, and a limited set of relevant variables. As a result, the problem scope is clearly defined and causal relationships between variables are understood (Reed, 2016). In managerial contexts, well-structured problems are usually optimization projects, like cost reduction or process improvements.

### *Ill-structured problems*

In contrast, an ill-structured problem is defined by ambiguous or evolving objectives, with information about the problem being incomplete or constraints being conflicting. There is no single correct solution, rather, the problem is open-ended, with multiple valid answers and trade-offs between them (Reed, 2016). Usually, ill structured problems arise in strategic or market-related contexts like redesigning an organization or responding to new external threats.

### *Wicked-structured problems*

A wicked problem is known for even greater levels of ambiguity and heavily contested problem definitions. There is no singular agreed-upon objective and therefore it lacks an inherent logic that signals when it is solved. Solutions in general are neither correct nor incorrect but rather better or worse with an unlimited number of possible options (Rittel & Webber, 1973). In reality, wicked problems occur in long-term strategic transformations like sustainability initiatives or adaption to a regulatory change.

### 2.5.3 Effective Problem Structuring

Given that managerial problems differ substantially in their level of structure and complexity of the solution, effective problem structuring does not follow a universal blueprint. In fact, it requires a specific solution for its problem type to effectively support managers. Accordingly, it is necessary to define clear evaluation dimensions that allow for a systematic assessment of how well large language models support managerial problem structuring.

### 2.5.4 Evaluation Dimensions

Although managerial problems vary, leaders generally rely on a consistent set of quality criteria when assessing whether a problem has been structured effectively.

#### *Logical Coherence*

Logical coherence refers to the internal consistency and clarity of the problem structure. This helps managers quickly grasp the structure of the output, as objectives, identified causes, and analytical steps follow a clear and traceable line of reasoning without contradictions. Literature suggests that many decisions are due to unclear framing rather than flawed execution (Hammond, Keeney, & Raiffa, 1998).

#### *Actionability*

Actionability ensures that managers can move from broader analysis to concrete steps and don't get lost in theoretical frameworks. Setting clear objectives, defining responsibilities, and tracking progress, helps managers to take action and ensure that decisions lead to tangible results (Boardwise, 2025).

#### *Comprehensiveness (MECE)*

Comprehensiveness describes whether a problem structure covers all relevant points without creating unnecessary overlap. Following McKinsey's MECE principle, a comprehensive

structure ensures a sufficient coverage of all important pain points while being analytical and organized (Rasiel, 1999).

### *Depth of insight*

Depth of insight gives the problem structure the necessary detail to be useful. It ensures that the output does not remain on the surface but digs deeper into underlying root causes of the problem, uncovering interdependencies and drivers critical for managerial interpretation (Branch, 2023).

## 2.5.5 Managerial Value of AI-Augmented Problem Structuring

Based on the demonstrated effectiveness of human–AI collaboration thus far, it is reasonable to assume that it will also prove valuable in supporting managers during the problem-structuring phase. In particular, large language models can relieve managers of common challenges when first framing complex problems. It can help overcome blank-page syndrome by generating initial ideas and alternative perspectives. It is also helpful when dealing with information overload by synthesizing scattered, unstructured information to support decision-making (Yun et al., 2025).

## 2.6 Research Gap

### 2.6.1 Purpose and scope of the literature review

This literature review establishes the theoretical foundation for examining how prompt engineering strategies influence the quality of generated problem structuring outputs by integrating three streams of research.

First it narrows the broad topic of artificial intelligence down to large language models and their role in managerial workflows. It breaks down how human-AI collaboration can benefit the decision-making process of leaders, emphasizing that outcomes depend strongly on how managers use this technology to their advantage.

Second, the review connects this research to literature on prompt engineering as the interface through which users interact with the large language models, highlighting evidence that prompt design has a significant impact on the model outputs.

Third, the chapter places prompt engineering within the existing literature on problem structuring and decision making. It states the importance of problem structuring as a critical phase at the beginning of managerial work.

This structure reveals a gap in the literature. Prior research focuses primarily on the benefits of generative AI for managerial decision making or the influence of prompt design on the model outputs but fails to systematically analyze differences in the quality of AI supported problem structuring. Moreover, existing studies do not compare effectiveness across different types of managerial problems with varying levels of structure and uncertainty. This gap provides the starting point for the research question and empirical design of this study.

## 2.6.2 Relevance and study positioning

This thesis is important because it addresses a critical phase of managerial decision making. The structure of a complex business problem strongly influences downstream analysis and strategic outcomes. Significant costs and inefficiencies can arise when this stage is poorly executed. As large language models are increasingly used to support managerial work, uncertainty about how prompting strategies shape the quality of AI supported problem structuring limits their effective use across industries.

This study addresses this gap by systematically comparing different prompting strategies in terms of their impact on the quality of AI generated problem structuring outputs. To increase applicability, it compares these effects across different types of managerial problems characterized by varying levels of structure and uncertainty. Thereby, providing evidence on how and under which conditions prompting strategies can support managerial problem structuring and directly motivating the research question.

## 3. Methodology

### 3.1 Research Design

To determine how different prompting strategies influence the quality of LLM-generated problem-structuring outputs, this thesis used a controlled experimental design. In order to achieve a more holistic picture on the LLM-generated output, this study followed a mixed-methods approach (Creswell, 2018). It combined a quantitative analysis in which experts rated the outputs for statistical comparison, with a qualitative analysis that examined reasoning patterns and structure in more detail. For the quantitative analysis, the prompt strategy served as the independent variable while the dependent variable was the quality of the generated output. The LLM version, temperature, token limit, and the respective cases were held constant as the control variables to ensure identical testing conditions across all prompts. The experiment followed a mixed design, with prompting strategy varied within evaluators and case type varied between evaluators. That allowed direct comparisons across prompts under the same condition.

The cases themselves were real consulting examples published by established firms as best practices for promotional purposes. The main goal of this experiment was to show how structured prompting can support better strategic reasoning and decision quality.

### 3.2 Experimental Setup

Before setting up the prompt engineering environment, three real consulting cases were selected, adapted, and used as inputs for the experiment to ensure realistic and context-driven problem situations. Each case was selected to represent a problem type defined in the literature, allowing an assessment of whether different problem types are better suited for AI supported problem structuring.

The well-structured problem was about improving a plastics producer's process to reduce its acetic acid consumption across several sites. While the ill-structured problem involved an airline which had to renegotiate major contracts and rebuild procurement capabilities after Covid-driven disruptions and collapsing travel demand. For the wicked problem, the

experiment focused on an oil and gas company that was struggling to build a clean energy business that delivered strong financial returns while reducing greenhouse gas emissions.

Any identifying details were removed from the prompt to prevent the model from relying on prior knowledge as well as remove familiarity bias of the evaluators. While some degree of situational steering was unavoidable, because every piece of information given to the LLM could influence the output, the experiment eliminated any skewness by providing identical information to each prompt strategy.

Each case was tested with all three prompting strategies in a separate, fresh model session so that no memory carryover effects could occur. In order to keep the LLM settings (model version, temperature, token limit, and system configuration) constant, all outputs were generated using the OpenAI API. Using deterministic settings (temperature = 0.0, top-p = 1.0) ensured that every prompt–case combination produced the exact same output every time, eliminating model randomness. This approach guaranteed a complete dataset covering every prompt condition across all cases under a controlled environment.

### 3.3 Prompt Design

Each of the three prompt strategies was framed through a dedicated prompt template, using the same case description to make the experiment not only transparent and reproducible but also comparable across all conditions. The templates differed only in structural guidance, while the case content stayed the same.

Baseline prompts provided only the case text without any additional structure. Instruction prompts added role guidance, format expectations, and output constraints. Decomposition prompts broke the task into sequential steps aligned with typical problem-structuring logic.

All prompts were kept consistent in tone and length, differing only in the structural constraints relevant to their respective strategy.

### 3.4 Operationalization of Prompting Strategies

Each prompt strategy followed a clear prompt format applied across all cases so they can be compared in a controlled and reproducible manner. As a result, any differences in output quality can be attributed to the prompting strategy itself rather than variations in task framing.

#### *Baseline Prompting*

For baseline prompting, only the case description with a minimal task instruction was provided. No role guidance, no other instructions, and no structural or formatting cues were prompted. This Output serves as the baseline condition to observe unguided model behavior.

#### *Instruction Prompting*

For instruction prompting, a standardized instruction block was added above the case description. This includes the core question, the number of issue areas and sub-questions, structural instructions, format expectations as well as a specific role definition.

#### *Decomposition Prompting*

This prompt directly builds on the instruction prompting input but adds an internal decomposition of the case. It instructs the LLM to break the case into high-level components, analyze them and determine the data required for each component. The components are placed into a logical sequence and grouped into four phases. The internal reasoning and decomposition steps were hidden from the final answer, so the input remains clean and comparable.

### 3.5 Prompt–Problem Matrix

Each of the three prompt strategies is tested across all three problem types, producing a fixed set of nine LLM-generated analytical outputs. These unique combinations of prompting strategy and case get assigned to a unique identifier to help with consistency and clarity across the quantitative as well as qualitative analysis. The table below presents an overview of all prompt–problem combinations and their respective output identifiers, as well as case abbreviations. Full prompt templates, inputs, and cases for each matrix entry are provided in the appendix.

| <b>Case</b> | <b>Problem type</b> | <b>Prompt strategy</b> | <b>Prompt ID</b> | <b>Output ID</b> |
|-------------|---------------------|------------------------|------------------|------------------|
| <b>A</b>    | Well-structured     | Baseline               | P01              | O01              |
| <b>A</b>    | Well-structured     | Instruction            | P02              | O02              |
| <b>A</b>    | Well-structured     | Decomposition          | P03              | O03              |
| <b>B</b>    | Ill-structured      | Baseline               | P04              | O04              |
| <b>B</b>    | Ill-structured      | Instruction            | P05              | O05              |
| <b>B</b>    | Ill-structured      | Decomposition          | P06              | O06              |
| <b>C</b>    | Wicked              | Baseline               | P07              | O07              |
| <b>C</b>    | Wicked              | Instruction            | P08              | O08              |
| <b>C</b>    | Wicked              | Decomposition          | P09              | O09              |

### 3.6 Evaluation Framework

Each output, corresponding to one prompt–case combination, was evaluated by a small panel of 41 people, consisting of both experienced domain experts and young professionals. By including both senior and junior evaluators, the experiment reflected the mixed-experience

teams that typically work on strategic problem structuring. The evaluation did not assess factual accuracy or client-specific realism but focused on the quality of the problem structure. To achieve comparable results, a standardized rating framework was used to evaluate the outputs across four core criteria: logical coherence, actionability, comprehensiveness (MECE), and clarity of communication. While each dimension was rated on a 5-point Likert scale (1 = very low, 5 = very high) to allow quantitative comparison across prompt conditions. In order to avoid order effects or familiarity bias, all outputs were anonymized and randomized before evaluation. All evaluators agreed to participate in this study and remained anonymous. The complete evaluation rubric with detailed dimension descriptions is provided in the appendix.

### 3.7 Data Analysis

For the quantitative analysis, ratings from all evaluators were compiled for each prompt-case combination. To summarize how each prompting strategy performs across all four evaluation dimensions, we used descriptive statistics (mean scores and variance). The resulting patterns in the aggregated scores were analyzed and compared in order to evaluate the differences between the prompting strategies. The test also examined how strategy performance varied across individual cases to identify whether certain prompting strategies were better suited to specific problem types.

To complement the quantitative analysis, a qualitative analysis of the output was conducted. This analysis followed a similar structure than the experiment, by focusing on the same four dimensions, to further contextualize and explain the quantitative patterns. It provided further insights on structural characteristics, strengths, and limitations across prompting strategies.

These integrated details provided a clear comparison of strengths and weaknesses of each prompt-case combination and gave insight into how different prompting strategies may support managerial problem structuring. The full analysis, including numerical results, distributions, and coded themes, is presented in the dedicated analysis chapter.

### 3.8 Validity and Reliability

To guarantee reliability across the experiment it was crucial to remove any randomness from the model outputs and ensure consistent operationalization of the prompting strategies. For that reason, it was important to keep model settings deterministic (temperature = 0.0, top-p = 1.0) and the prompt templates fixed. Moreover, a standardized rating rubric was used by all evaluators to support consistent scoring by all participants, while all outputs were anonymized and randomized before evaluation to reduce evaluator bias.

The validity of the experiment was achieved by keeping consistency across model version, temperature, token limit, case content and system configuration. As a result, differences in model output within each case can be attributed to the prompting strategy.

Using a within-conditions design, where each prompting strategy was applied to the same case. This increased comparability across conditions, while randomization prevented ordering and familiarity effects. Each evaluation criteria was grounded in the literature on effective problem structuring, so the rating system directly measured the quality of the structure the LLM suggests. Senior and junior evaluators improved ecological validity by reflecting how strategic problems are typically assessed in real organizational environments. Using real consulting cases further increased the realism and applicability of the findings. Inter-rater reliability statistics are not reported because evaluators did not rate the same set of outputs, making measures such as ICC inappropriate.

## 4. Quantitative Analysis

The quantitative analysis was based on an online survey of 41 evaluators. Each participant was randomly assigned to one of three problem cases, each one representing a different level of problem structure (well-structured, ill-structured, and wicked-structured), to assess three AI-generated problem-structuring outputs, associated with different prompting strategies (baseline prompting, instruction prompting, and decomposition prompting).

The rating-based analysis resulted in 492 individual observations using a five-point Likert-scale. As a result of evaluators contributing multiple ratings, controlling for systematic differences in individual rating behavior in the following regression analysis is necessary.

In addition to the rating-based analysis, evaluators were asked to rank each problem structuring output from best to worst, resulting in 123 observations, corresponding to three rankings per evaluator.

The analysis combined both rating and ranking data, allowing for inferential analysis of prompting strategy effects and complementary robustness checks based on relative preferences.

### 4.1 Descriptive Statistics

The mean evaluation score across all 492 observations was 3.62 with a standard deviation of 1.02, indicating moderate dispersion around the midpoint of the Likert-scale.

Aggregating by prompting strategy, the data showed higher average ratings for instruction and decomposition prompting compared to the baseline, suggesting that prompting strategies which force more structure correlate with more favorable results.

Across evaluation dimensions, logical coherence scored a higher average rating relative to other dimensions, while differences in comprehensiveness and depth are less pronounced. This indicates that differences in structural clarity were more strongly perceived by evaluators than other dimensions.

Descriptive statistics suggest only limited differences across case types, although inferential analysis reveals a statistically significant negative effect for the ill structured case.

## 4.2 Model specification and estimation approach

For examining the different effects of prompting strategies on evaluation ratings, an ordinary least squares (OLS) regression framework was used. The evaluation rating functioned as the dependent variable while the prompting strategy, operationalized as a categorical variable was used as the independent variable.

Additionally, the model accounted for context by including case type and different evaluation dimensions (logical coherence, comprehensiveness, depth, and actionability). Interaction terms between strategy and case types were included to examine the difference in effectiveness across prompting strategies and problem types.

To combat a violation of the OLS independence assumption, the model included rater fixed effects, ensuring that estimated strategy effects were identified by comparing how the same evaluator rates different strategies. Robust standard errors are used to account for heteroskedasticity and conservative inference across the experiment.

### 4.3 Regression Results

| Variable                              | Coefficient ( $\beta$ ) | Robust Std. Error | p-value |
|---------------------------------------|-------------------------|-------------------|---------|
| Instruction prompting                 | 0.346                   | 0.162             | 0.033   |
| Decomposition prompting               | 0.808                   | 0.154             | < 0.001 |
| Ill-structured case                   | -1.192                  | 0.398             | 0.003   |
| Wicked-structured case                | -0.225                  | 0.328             | 0.494   |
| Comprehensiveness                     | 0.358                   | 0.113             | 0.002   |
| Depth                                 | 0.000                   | 0.119             | 1.000   |
| Logical coherence                     | 0.642                   | 0.108             | < 0.001 |
| Instruction $\times$ Ill-structured   | 0.300                   | 0.256             | 0.243   |
| Decomposition $\times$ Ill-structured | 0.026                   | 0.268             | 0.924   |
| Instruction $\times$ Wicked           | 0.404                   | 0.220             | 0.067   |
| Decomposition $\times$ Wicked         | 0.020                   | 0.214             | 0.924   |
| Constant                              | 3.365                   | 0.254             | < 0.001 |

The table above shows the main regression results of the relationship between prompting strategy and evaluation ratings, taking into account case type, evaluation dimension and rater fixed effects. The model explained a substantial share of the variation in evaluation ratings, with an  $R^2$  of 0.432 and an adjusted  $R^2$  of 0.369. The model overall was statistically significant ( $F = 6.869$ ,  $p < 0.001$ ).

### *Effects of prompting strategies and case heterogeneity*

The results showed clear variation depending on the type of prompt strategy used. Baseline prompting led to significantly lower evaluation outcomes compared to instruction prompting ( $\beta = 0.35$ ,  $p = 0.033$ ). Decomposition prompting showed an even greater positive effect ( $\beta = 0.81$ ,  $p < 0.001$ ), suggesting that prompts which more strongly constrain, and guide output improve the quality of AI-generated problem structures.

Interaction terms display limited evidence of heterogeneity across problem types. None of the interactions between prompting strategy and case type were statistically significant at conventional levels. The only marginal pattern appeared for instruction prompting in the wicked-structured case ( $\beta = 0.40$ ,  $p = 0.067$ ). Overall, the effects of prompting strategies appear largely stable across cases.

### *Evaluation dimensions*

Analyzing the evaluation dimensions, ratings for logical coherence were on average higher than ratings for the reference evaluation dimension ( $\beta = 0.64$ ,  $p < 0.001$ ). In addition, ratings for comprehensiveness were also significantly higher on average ( $\beta = 0.36$ ,  $p = 0.002$ ), while ratings for depth did not differ significantly from the reference category.

## 4.4 Ranking Analysis

For more robust results a ranking of the generated outputs from best to worst was included. In contrast to the rating task, this captured the ranking relative preferences between outputs and were therefore analyzed descriptively.

*Descriptive ranking results*

| <b>Case Type</b>         | <b>Prompting Strategy</b> | <b>Mean Rank</b> | <b>Median Rank</b> |
|--------------------------|---------------------------|------------------|--------------------|
| <b>Well-structured</b>   | Baseline                  | 2.54             | 3                  |
| <b>Well-structured</b>   | Instruction               | 1.85             | 2                  |
| <b>Well-structured</b>   | Decomposition             | 1.62             | 1                  |
| <b>Ill-structured</b>    | Baseline                  | 2.67             | 3                  |
| <b>Ill-structured</b>    | Instruction               | 1.83             | 2                  |
| <b>Ill-structured</b>    | Decomposition             | 1.50             | 1                  |
| <b>Wicked-structured</b> | Baseline                  | 2.81             | 3                  |
| <b>Wicked-structured</b> | Instruction               | 2.00             | 2                  |
| <b>Wicked-structured</b> | Decomposition             | 1.19             | 1                  |

The Table above shows the mean and median ranks of each prompting strategy by case. Lower values indicate better rankings. Across all cases, baseline prompting was consistently ranked lowest, with mean ranks ranging from 2.54 to 2.81. In the well-structured case, decomposition prompting ranked highest (mean = 1.62), followed closely by instruction prompting (mean = 1.85). For more complex cases (ill-structured and wicked-structured), decomposition prompting was also the best ranked prompting strategy (mean = 1.50 and 1.19, respectively).

Ranking differences were statistically significant for all three cases (ill-structured:  $p = 0.0018$ , well-structured:  $p = 0.0125$ , wicked-structured:  $p < 0.001$ ), based on a Friedman test with the null hypothesis that all prompting strategies receive equal mean ranks within each case.

### *Interpretation and consistency with rating results*

The ranking results were consistent with the regression analysis. In the well-structured case, instruction and decomposition prompting were narrowly ranked first and second, while decomposition prompting clearly outperformed instruction prompting in the ill-structured and wicked-structured cases. Baseline prompting consistently achieved the lowest rankings. Ranking differences were statistically significant across all three problem types, providing additional support for systematic differences between prompting strategies.

Overall, the quantitative analysis showed that prompting strategies primarily differ in their ability to induce structured outputs, with these differences becoming more pronounced for more complex problems.

## 5. Qualitative Analysis

The qualitative analysis was an author-led evaluation of the LLM-generated problem-structuring outputs. It built on the same framework used in the quantitative analysis, assessing across the four evaluation dimensions: Logical Coherence, Comprehensiveness (MECE), Actionability and Depth of Insight. The benefit of this approach was that it allowed for conceptual consistency across evaluations to complement and contextualize the quantitative rating patterns. The goal was not to re-score the outputs but to analyze them comparatively across different prompting strategies and case types by systematically examining recurring structural features within each evaluation dimension. Case A was used as an analytical anchor to observe strategy-level patterns, which were subsequently examined for robustness across the ill-structured and wicked cases.

### 5.1 Logical Coherence

This section examines how clearly, and consistently analytical steps are linked to one another within each problem structure.

While logical coherence read well across all strategies, the baseline output relied on indirect transitions between steps that the manager was expected to assume rather than on explicitly stated dependencies. The baseline structure presumed shared managerial intuition about objectives, success criteria, and how analytical findings translated into decisions, which resulted in a coherent linear flow but weak traceability across stages.

Instruction prompting improved with this by specifying data requirements and expected outputs. The resulting artifacts were handover points that help the step-by-step logic and reduce ambiguity in how analytical stages are connected. This led to clearer traceability between data preparation, driver identification, and the translation of findings into improvement levers.

Decomposition prompting showed the strongest logical dependency structure in Case A, by introducing specific benchmarks and driver trees to provide a pathway from diagnosis to implementation. By conditioning each step on the output of the previous step, the approach minimized assumptions from leaders and reduced the risk of logical gaps.

Baseline prompting therefore produced a coherent linear flow for Case A but lacked clarity of dependencies and intermediate handoffs. At this stage of the analysis, these limitations remained manageable because Case A represented a well-structured problem with relatively clear constraints and objectives.

These mechanisms also held true across cases but became more consequential as problem ambiguity increased. In Case B, which involved multiple competing objectives, the explicit outputs as a result of instruction prompting had a greater impact on coherence than in the previous case, because the next logical steps are more complicated to guess. This became even more noticeable for Case C, where the logical path was particularly unclear. The clear KPIs made it necessary for the LLM to define the objectives upfront, since they were not clearly outlined and not easily derived from the overarching goal. The explicit dependency chains introduced by decomposition prompting provided a necessary structure.

In conclusion, differences in logical coherence across strategies did not lie in other analytical steps but rather in the extent to which dependencies and handoffs were made explicit. Increasing problem ambiguity benefits from prompting strategies that enforce a more rigorous structure, while strategies relying on implicit logic become increasingly fragile.

## 5.2 Comprehensiveness (MECE)

This section assesses how completely and non-overlapping the problem space is covered by the proposed analytical structure.

The baseline prompt for Case A covered the core end-to-end logic of a typical improvement cycle. It included framing, data preparation, driver analysis, and testing levers which are crucial steps for a holistic problem structure. However, it remained surface-level by ignoring important sub-branches like governance, measurement cadence, or sustainment mechanisms. These omissions indicated incomplete coverage when it comes to comprehensiveness,

Instruction prompting on the other hand, expanded its coverage by specifying data needed and explicit outputs for each step, which made the resulting structure seem more complete. It still bundled a few points together under the umbrella term „key drivers“, including process,

equipment, or other variables. This typically weakened perceived MECE separation, even with complete coverage, by not covering structurally distinct problem dimensions.

Decomposition prompting introduced artifacts which force explicit separation of sub-problems like benchmarking or process mapping thereby reducing overlap between steps and exposing gaps in the plan.

Looking at Case B and C, the same pattern persisted. The high-level plan from baseline prompting covered the major bases but missed separating broader points like “internal capability fit“ in Case C which could lead to missing branches for a wicked strategy problem.

Overall, decomposition prompting remained most comprehensive by separating the overarching drivers into separate, not overlapping sub-problems through artifacts and outputs. Instruction and especially baseline prompting remained more surface level, but some overlap or gaps might be introduced through broader steps.

### 5.3 Actionability

This section evaluates how clearly the problem structure indicates what a manager should do next at each stage.

In terms of actionability, baseline prompting for the Case A was extremely limited. It might suggest reasonable step progression, but the plan did not clearly specify what a manager should do after completing one step. It would require significant interpretation and critical thinking to translate the structure into concrete actions.

While instruction prompting improved actionability by stating explicit outputs for each step and thereby reducing ambiguity to some extent, the specific actions required to generate these outputs remain implicit.

Decomposition prompting however, provided very clear instruction. It defined explicit “brief action[s]“, providing clear instructions on what to do next and linking the action directly to a defined output. This minimized the interpretive effort and clearly provided the strongest form of actionability.

Well formulated analytical sequence and logical coherence (cf. 6.1) can also help with actionability by reducing uncertainty about progression. For specific actions, it still relied heavily on implicit reasoning. In Case A, decomposition prompting was the only prompting strategy which clearly defined actions for managers to adopt.

When extending the analysis to Case B, actionability became more sensitive to ambiguity in the problem set. Baseline prompting still outlined relevant workstreams but left managers with multiple plausible next actions at each stage, making actionability worse. Interestingly, instruction prompting also introduced “brief action[s]“, for managers to strictly follow. However, the lack of clear dependencies between steps could still lead to confusion on which action to take first.

Decomposition prompting had a clearer sequencing of analytical steps. The ordering of actions made dependencies between steps more visible and reduces uncertainty, which improves actionability.

For the wicked problem in Case C, none of the prompting strategies provided explicit next-step action instructions. The absence of stable success criteria and the presence of evolving constraints might make prescriptive next-step instructions risky, leading to even structured prompts to prioritize sensemaking over explicit action guidance. Regardless, the highest level of ambiguity can still be found with baseline prompting. In contrast, instruction and decomposition prompting partially compensated for this by providing clearer structural ordering and logical coherence, which established a path forward even though concrete next actions remain implicit.

Across cases, actionability increasingly relied on structures that reduce interpretive burden. Prompting strategies that explicitly state what to do next or at the very least establish an implicit step-by-step framework, provide stronger actionability.

## 5.4 Depth of Insight

This section analyzes how explicitly underlying reasoning, trade-offs, and causal logic are articulated within the problem structure.

Looking at the depth of insight in Case A, the baseline prompt remained quite generic and did not go into detail. It mentioned plausible hypotheses and methods but did not translate them further into a hierarchy or causal chain.

Instruction prompting goes beyond “analyze drivers“ by specifying how performance should be normalized, compared across sites, and quantified in terms of impact.

Decomposition prompting showed the strongest depth of insight. By separating statistical identification of consumption drivers from the following technical validation, for example, it demonstrated deeper domain understanding, as it avoids treating all data correlations as real causes.

The observations for baseline prompting held true across problem types by sticking to broad coverage and generic techniques to structure a problem. Instruction prompting increased the depth of insight through domain-specific concepts like the synergy potential across pillars in Case C or supplier financial health and capacity in Case B. This provided a more decision-relevant structure than baseline prompting.

Decomposition prompting produced the deepest level of insight. It made trade-offs between cost, risk, and service for the ill-structured problem very clear and linked them to a longer-term change-management plan. In Case C, it tied value-creation to profit pools and differentiation as well as specifies governance and risk-mitigation mechanisms. This showed a higher level of reasoning than both previous strategies.

However, it is worth mentioning that the depth of insight was inherently more interpretive than the other dimensions. Higher depth of insight did not necessarily mean a high-quality of problem structure, as some managers benefit or prefer a higher-level problem frame.

## 6. Discussion

The literature review identified a lack of empirical evidence on how prompt engineering strategies influence the quality of AI-supported problem structuring, especially across different types of managerial problems. Prior research so far has primarily discussed general benefits of generative AI in managerial workflows or analyzed the effect of prompt design on the model outputs.

The findings of this study close this gap by showing a systematic influence of prompting strategies on the quality of the problem structure. Instruction prompting and decomposition prompting consistently outperformed baseline prompting across all tested problem types. These results showed that prompt design affects problem structuring quality beyond stylistic differences and independently of problem type.

### 6.1 Why Structured Prompting Works

The results of the analysis suggested that prompting strategies in the context of strategic problem structuring primarily differ in how explicitly they enforce analytical structure. Evaluators tend to score a problem structure higher if it specified intermediate outputs, dependencies between steps, and expected immediate actions, instead of the manager having to rely on his or her implicit reasoning.

Logical coherence appears to be most strongly associated with the overall evaluation ratings. The qualitative analysis explains this relationship by showing that explicit dependency chains reduce ambiguity about potential next steps. Moreover, a logical coherent plan can positively influence other dimensions like actionability which also results in a higher score.

As the problems became more complex and the goals more ambiguous, this effect increased. Unclear goals, constraints, and causal relationships make it increasingly difficult for managers to construct robust problem structures. Prompts that force the LLM to organize reasoning in a way that is easier to evaluate, communicate, and act upon help stabilize the analytical process by externalizing assumptions and sequencing. More guiding prompts therefore improved the quality of the resulting problem structure.

## 6.2 Practical Implications and Boundary Conditions

Linking the findings of the quantitative and the qualitative analysis, it is safe to assume that prompt engineering is more than a just technical task but rather a design choice. For very complex and ambiguous tasks it might be worth engineering a prompt that explicitly specifies actions, intermediate outputs, and dependencies. In contrast, for simpler, more straight forward problems or just a quick overview, lighter prompting may be sufficient.

The implications derived from this study are based on a controlled experimental setting and should therefore be interpreted with care. The results are merely indicators instead of universally applicable. Within these boundaries, this analysis shows potential of prompt strategies to support managers at an early and critical stage of decision making.

## 7. Conclusion

The goal of this thesis was to examine how different prompt engineering strategies influence the quality of AI-supported problem structuring in managerial decision-making contexts. Through a combination of a controlled experiment with quantitative ratings, ranking data, and qualitative analysis across a variety of problem types, the study provides systematic empirical evidence on an area that has not been the topic of research to that extent.

The results show a clear and consistent pattern. Prompting strategies which guide the output in the context of problem structuring more, outperformed minimal baseline prompting consistently. Instruction prompting improved problem-structuring quality by making expected outputs and analytical steps more explicit, while decomposition prompting delivered the strongest effects by clearly separating sub-problems and recommending immediate actions to minimize implicit interpretation by managers.

Across evaluation dimensions, logical coherence was the most important driver of problem structure quality. Outputs that clearly articulated dependencies between steps and made intermediate handoffs explicit were consistently rated higher. Depth of insight on the other

hand, did not show as consistent effects and appeared to be context dependent. The qualitative analysis confirmed these findings, especially as problem complexity increased.

Overall, the findings demonstrate that prompt engineering is not purely a technical task but a managerial design choice with significant implications for decision quality. Well-designed prompts can support managers at an early and critical stage of problem solving by stabilizing reasoning, improving communicability, and enhancing actionability. This study contributes empirical evidence to the literature on AI-supported managerial work and highlights prompt design as a lever for improving strategic problem structuring.

## 8. Limitations and Future Research

This study, like any other, is subject to several limitations that should be considered when interpreting its findings. Every design choice made to ensure experimental control and comparability introduces limitations that are unavoidable.

First, the number of tested cases was limited, which made it difficult to generalize across industries and organizations. This was especially the case, because every problem representation had to be reduced and simplified, which further decreased the complexity, dynamics and uncertainty of the real world.

Second, the scope of this experiment was inherently constrained by its prompt design. The study examined a set of predefined prompting strategies and relied on a single prompt per strategy and case. While this improved internal validity and comparability, it did not fully reflect the way many people use large language models. By not allowing the evaluators to frame the prompts themselves based on their preference or adapt them iteratively (Chain-of-thought-prompting), the natural way humans and language models interact was not captured. Moreover, complete neutrality in prompt design could not be achieved which always introduces some sort of bias. Any specific wording or structural choice necessarily influenced the generated output beyond the abstract prompting strategy itself and could never be entirely neutralized.

Third, the use of a single AI model raised questions about whether the observed patterns could transfer to other models, providers or model versions. Given the rapidly evolving nature of

large language models there are difficulties associated with making broad predictions on how the results might change in the future. The findings should therefore be interpreted as model- and time-specific, while being an indicator at a conceptual level.

Finally, there were certain limitations that also arise from the use of Likert-scale ratings. It restricts measurement granularity and introduces hard scale limits which can obscure small differences between outputs. This is particularly the case because all outputs did not differ strongly from one other, so nuanced differences could get lost in this measurement approach. In addition, the use of both rating-based and ranking-based evaluations shows that using different measurement approaches may capture different aspects of the respective output. A regression analysis highlights the absolute differences in output quality while the rankings reflect the relative performance, which introduces a measurement-related limitation.

These limitations highlight the strategic boundaries of the study and suggest that the findings should be interpreted with appropriate caution. They also show potential for future research that could broaden the scope and robustness of this work. By examining a larger and more diverse set of cases, the application horizon of these insights would be greatly increased. Furthermore, further study could explore a broader, multi-level form of prompting strategies, for a more realistic reflection of human-AI augmented problem structuring.

Overall, this study provided a controlled and systematic starting point for examining how different prompting strategies shape AI-supported problem structuring in realistic business cases, while highlighting clear directions for future research.

## 9. Bibliography

Eloundou, T., Manning, S., Mishkin, P., & Rock, D. (2023). GPTs are GPTs: An early look at the labor market impact potential of large language models (arXiv:2303.10130). arXiv.

<https://doi.org/10.48550/arXiv.2303.10130>

Dell'Acqua, F., Fuller, J., Kim, H., & Kleinbaum, A. M. (2023). Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality (Harvard Business School Technology & Operations Management Unit Working Paper No. 24-013). Harvard Business School.

<https://doi.org/10.2139/ssrn.4566546>

Fortune Business Insights. (2024, April). Prompt engineering market size, share, and industry analysis, by component (software and services), by technique (generated knowledge prompting, n-shot prompting, and others), by application (conversational AI, content generation, recommendation systems, and others), by industry (healthcare, BFSI, media and entertainment, retail, and others), and regional forecast, 2025–2032.

<https://www.fortunebusinessinsights.com/prompt-engineering-market-109382>

Tinguely, P. N., Lee, J., & He, V. F. (2023). Designing human resource management systems in the age of AI. *Journal of Organization Design*, 12(1), 1–7. <https://doi.org/10.1007/s41469-023-00138-x>

Coursera. (2024). What is artificial intelligence? Coursera Articles. Retrieved October 21, 2025, from <https://www.coursera.org/articles/what-is-artificial-intelligence>

Quiroz-Vázquez, C., & Goodwin, M. (n.d.). Artificial intelligence in business. IBM Think. Retrieved January 3, 2026, from <https://www.ibm.com/think/topics/artificial-intelligence-business>

McKinsey & Company. (2024a). What is AI (artificial intelligence)? McKinsey Explainers. Retrieved October 21, 2025, from <https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-ai>

McKinsey & Company. (2024b). The state of AI in early 2024: Gen AI adoption spikes and starts to generate value. McKinsey Insights. Retrieved October 21, 2025, from <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-2024>

Han, S., Wang, M., Zhang, J., Li, D., & Duan, J. (2024). A review of large language models: Fundamental architectures, key technological evolutions, interdisciplinary technologies integration, optimization and compression techniques, applications, and challenges. *Electronics*, 13(24), 5040. <https://doi.org/10.3390/electronics13245040>

Stryker, C. (n.d.). What is a domain-specific LLM? IBM Think. Retrieved January 3, 2026, from <https://www.ibm.com/think/topics/domain-specific-llm>

Bergmann, D. (n.d.). What is instruction tuning? IBM Think. Retrieved January 3, 2026, from <https://www.ibm.com/think/topics/instruction-tuning>

Varughese, J. (n.d.). What is a multimodal LLM (MLLM)? IBM Think. Retrieved January 3, 2026, from <https://www.ibm.com/think/topics/multimodal-llm>

McKinsey & Company. (2024c). The economic potential of generative AI: The next productivity frontier. McKinsey Insights. Retrieved October 21, 2025, from <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier>

Kolbjørnsrud, V. (2024). Designing the intelligent organization: Six principles for human–AI collaboration. *California Management Review*. Retrieved October 29, 2025, from <https://cmr.berkeley.edu/2024/02/66-2-designing-the-intelligent-organization-six-principles-for-human-ai-collaboration/>

Berretta, S., Tausch, A., Ontrup, G., Gilles, B., Peifer, C., & Kluge, A. (2023). Defining human–AI teaming the human-centered way: A scoping review and network analysis. *Frontiers in Artificial Intelligence*, 6, [1250725](https://doi.org/10.3389/frai.2023.1250725). Retrieved October 29, 2025, from <https://www.frontiersin.org/articles/10.3389/frai.2023.1250725/full>

Cappelli, P., Tambe, P. (Sonny), & Yakubovich, V. (2024, March 4). Will large language models really change how work is done? *MIT Sloan Management Review*. Retrieved March 4, 2024, from <https://sloanreview.mit.edu/article/will-large-language-models-really-change-how-work-is-done/>

Vaccaro, M., Almaatouq, A., & Malone, T. (2024). When combinations of humans and AI are useful: A systematic review and meta-analysis. *Nature Human Behaviour*, 8, 2293–2303.

<https://doi.org/10.1038/s41562-024-02024-1>

Przegalinska, A., Triantoro, T., Kovbasiuk, A., Ciechanowski, L., Freeman, R. B., & Sowa, K. (2025). Collaborative AI in the workplace: Enhancing organizational performance through resource-based and task–technology fit perspectives. *International Journal of Information Management*, 81, 102853. <https://doi.org/10.1016/j.ijinfomgt.2024.102853>

Boston Consulting Group. (2024). Organizations combining organizational learning and AI-specific learning see greater success. Press release. Retrieved October 29, 2025, from

<https://www.bcg.com/press/12november2024-organizational-learning-and-ai-specific-learning-managing-uncertainty>

Romeo, M., Karayannis, T., & Falco, J. (2025). Exploring automation bias in human–AI collaboration. *AI & Society*. Retrieved October 29, 2025, from

<https://link.springer.com/article/10.1007/s00146-025-02422-7>

Heyder, T. (2023). Ethical management of human–AI interaction: Theory and evidence. *AI & Ethics*, 3(2), 145–159. Retrieved October 29, 2025, from

<https://www.sciencedirect.com/science/article/pii/S0963868723000185>

Bankins, S. (2024). A multilevel review of artificial intelligence in organizations. *Journal of Organizational Behavior*, 45(1), 1–22. Retrieved October 29, 2025, from

<https://onlinelibrary.wiley.com/doi/10.1002/job.2735>

Pandit, B. (2024, August 22). Prompt engineering fundamentals: From zero to hero with prompt templates for different types of prompts. IBM Developer.

<https://developer.ibm.com/articles/awb-prompt-engineering-fundamentals/>

Google Cloud. (n.d.). Prompt engineering for AI guide. Retrieved October 29, 2025, from

<https://cloud.google.com/discover/what-is-prompt-engineering>

Sahoo, P., Singh, A. K., Saha, S., Jain, V., Mondal, S., & Chadha, A. (2024). A systematic survey of prompt engineering in large language models: Techniques and applications. arXiv preprint. Retrieved October 29, 2025, from <https://arxiv.org/abs/2402.07927>

Bonna, S., Huang, Y.-C., Novozhilova, E., Paik, S., Shan, Z., Feng, M., Gao, G., Tayal, Y., Kulkarni, R., Yu, J., Divekar, N., Ghadiyaram, D., Wijaya, D., & Betke, M. (2025). DebiasPI: Inference-time debiasing by prompt iteration of a text-to-image generative model. arXiv preprint. <https://doi.org/10.48550/arXiv.2501.18642>

Schulhoff, S. (2025). Technique #1: Instructions in prompts. Learn Prompting. Retrieved December 21, 2025, from <https://learnprompting.org/docs/basics/instructions>

Kuka, V. (2024, September 27). Advanced decomposition prompting techniques. Learn Prompting. <https://learnprompting.org/docs/advanced/decomposition/>

Brucks, M., & Toubia, O. (2025). Prompt architecture induces methodological artifacts in large language models. PLOS ONE. Retrieved October 30, 2025, from <https://journals.plos.org/plosone/article?id=10.1371%2Fjournal.pone.0319159>

Sun, Y., & Kok, S. (2025). Investigating the effects of cognitive biases in prompts on large language model outputs. arXiv preprint. Retrieved October 30, 2025, from <https://arxiv.org/abs/2506.12338>

Chadimová, M., Jurášek, E., & Kliegr, T. (2024). Meaningless is better: Hashing bias-inducing words in LLM prompts improves performance. arXiv preprint. Retrieved October 30, 2025, from <https://arxiv.org/pdf/2411.17304>

McKinsey & Company. (2024d). What is prompt engineering? McKinsey Explainers. Retrieved October 30, 2025, from <https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-prompt-engineering>

Barney, J. B., Ketchen, D. J., Jr., & Wright, M. (2021). Resource-based theory and the value creation framework. *Journal of Management*, 47(7), 1936–1955. <https://doi.org/10.1177/01492063211021655>

Anthropic. (2024). Prompt engineering for business performance. News release. Retrieved October 30, 2025, from <https://www.anthropic.com/news/prompt-engineering-for-business-performance>

Binder, J., & Watkins, M. D. (2024). To solve a tough problem, reframe it. *Harvard Business Review*, January–February 2024. Retrieved November 8, 2025, from <https://hbr.org/2024/01/to-solve-a-tough-problem-reframe-it>

Reed, S. K. (2016). The structure of ill-structured (and well-structured) problems revisited. *Educational Psychology Review*, 28(4), 691–716. <https://doi.org/10.1007/s10648-015-9343-1>

Rittel, H. W., & Webber, M. M. (1973). "Dilemmas in a General Theory of Planning." *Policy sciences*, 4(2), 155-169. <https://www.cc.gatech.edu/fac/ellendo/rittel/rittel-dilemma.pdf>.

Hammond, J. S., Keeney, R. L., & Raiffa, H. (1998). The hidden traps in decision making. *Harvard Business Review*, 76(5), 47–58. Retrieved November 8, 2025, from <https://hbr.org/1998/09/the-hidden-traps-in-decision-making-2>

Boardwise. (2025). What is an action plan and why is it essential for boardroom success? Boardwise Blog. Retrieved December 22, 2025, from <https://www.boardwise.io/en/blog/what-is-an-action-plan-and-why-is-it-essential-for-boardroom-success>

Rasiel, E. M. (1999). *The McKinsey way*. McGraw-Hill.

Branch, E. (2023, December 30). Mastering the 5 Whys: Unlocking deep insight. IENSTITU. Updated August 25, 2024. <https://www.ienstitu.com/en/blog/deep-insight>

Yun, B., Feng, D., Chen, A. S., Nikzad, A., & Salehi, N. (2025). Generative AI in Knowledge Work: Design Implications for Data Navigation and Decision-Making. *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '25)*, Yokohama, Japan. Association for Computing Machinery. <https://doi.org/10.1145/3706598.3713337>

# 10. Appendix

## Test-Cases

### Case A

#### **Advanced Analytics Helps a Plastic Manufacturer Optimize its Yield**

(Well-Structured problem)

##### *Original*

A leading producer of plastics sought our help improving its production processes across seven global sites. In particular, ChemicalsCo wanted to find ways to reduce its consumption of a key ingredient in many plastics. In just four weeks of analysis, we helped ChemicalsCo reduce consumption of the ingredient, acetic acid, by 20%

In the years leading up to our engagement, ChemicalCo's consumption of acetic acid had been high, and it had been one of the company's biggest variable costs. To help solve the problem, we assembled a joint team of consultants, data scientists, and process engineers.

Our team used a combination of machine learning techniques and engineering knowledge to distill the key factors impacting acetic acid loss. This allowed us train a simple, transparent algorithm to identify the factors determining periods of "low" and "high" consumption in ChemicalCo's processes. The insights from the model were immediately actionable for ChemicalsCo to reduce its ingredient and thereby bring more stability to production costs across its global footprint.

\* We take our clients' confidentiality seriously. While we've changed their names, the results are real.

##### *Standardized*

#### 1. Context

A leading plastics producer operates seven global production sites. Acetic acid is a key ingredient in many of its plastics manufacturing processes and represents one of the company's largest variable cost components.

#### 2. Core Problem

The company has experienced consistently high acetic acid consumption over several years and does not understand which factors within its processes are driving this elevated usage.

#### 3. Situation at Start

At the beginning of the project, the company only knows that acetic acid consumption has been high for an extended period and that the underlying reasons remain unclear. Consumption levels vary, but no explanation or systematic analysis has yet been carried out, and internal teams suspect that process-related factors may be contributing. A joint team of consultants, data scientists, and process engineers is planned to address the issue.

#### 4. Known Constraints

The issue must be examined across all seven global sites, and any understanding of improvement levers should preserve stable production while addressing significant variable costs linked to acetic acid.

#### 5. Objective

The company aims to identify the key factors that influence acetic acid consumption and determine where opportunities may exist to reduce usage.

## **Case B**

### **How an Airline Used a Global Stall to Give Procurement a Lift**

**(Ill-Structured problem)**

#### *Original*

A major APAC airline coming out of administration sought significant cost reductions through procurement savings and by building best-in-class procurement capabilities. We provided support in three ways: by assessing the entire spending base to estimate potential savings; by identifying urgent priorities given market conditions and “nonnegotiable” contract deadlines; and by mapping out existing capability gaps in the sourcing life cycle. AirlineCo had experienced substantial disruptions in the industry and supplier landscape due to the Covid-19 pandemic, prompting the administration team to reset all contracts. This reset necessitated renegotiating multiple high-value, long-term contracts with a relatively new procurement team. The pandemic also changed travel usage patterns due to reduced business travel amid safety concerns and organizational downsizing, decreased international flights, and other impacts, which further complicated procurement decisions.

We performed an in-depth diagnostic to assess historic spending across major procurement categories to identify cost levers, key contract terms, service-level agreements (SLAs), and key performance indicators (KPIs). We then benchmarked cost drivers and prioritized high-value categories that needed immediate action. Multiple categories—such as in-flight

catering, ground handling, fuel, maintenance, repair and operations (MRO) components, engine maintenance, and base maintenance—were flagged for deep analysis followed by the creation of a detailed roadmap and list of initiatives to meet the identified savings target. After completing the diagnostic, we provided full procurement process assistance for high-value categories through the development of sourcing and negotiation strategies, request for proposal (RFP) preparation and execution, evaluation of offers, and supplier selection. We also helped AirlineCo establish planning, supplier relationship management, and contract management capabilities including:

- integrating an annual procurement plan that encompasses refined cost categorization, spending baseline, and savings ambition;
- defining decision rights for the end-to-end procurement process;
- developing supplier selection methodology and shortlisting; and
- supporting a contract tracking tool and management plans for long-term vendors.

As a result, we identified and implemented 14–16% savings across categories, which was 50% higher than the outside-in targets. Our holistic assessment of AirlineCo unlocked critical savings across categories, reorganized its sourcing strategy, streamlined its vendor base, and enabled it to chart a path to becoming a world-class procurement organization.

### Standardized

#### 1. Context

A major APAC airline had gone through administration and was emerging from a period of severe industry and supplier disruptions caused by the Covid-19 pandemic. Travel usage patterns had shifted significantly due to reduced business travel, safety concerns, organizational downsizing, and decreased international flights. These changes affected demand, operations, and the supplier landscape.

#### 2. Core Problem

The airline needed to achieve significant cost reductions and rebuild strong procurement capabilities after the pandemic and administration process.

#### 3. Situation at Start

Following administration, all contracts had been reset, requiring the airline to renegotiate multiple high-value, long-term agreements with a relatively new procurement team. The company faced urgent timelines driven by market conditions and nonnegotiable contract deadlines. Major procurement categories such as in-flight catering, ground handling, fuel,

maintenance, repair and operations components, engine maintenance, and base maintenance were large cost drivers that required attention. The airline needed clarity on its spending base, contract terms, and procurement capability gaps.

#### 4. Known Constraints

Renegotiations had to occur under time pressure, influenced by market conditions, pandemic-related disruptions, and fixed contract deadlines. The procurement team was relatively new, and the airline operated across multiple high-value categories that had been reset during the administration process.

#### 5. Objective

The airline aimed to reduce procurement costs and build best-in-class procurement capabilities to support stable operations after administration.

## **Case C**

### **An Oil and Gas Leader Creates a New Clean Energy Business**

**(Wicked-Structured problem)**

#### *Original*

A leading oil and gas company wanted to create a clean energy solutions business. Within just nine months, we helped them design and launch CleanEnergyCo\*, an independent entity that promises to produce substantial financial returns while contributing to the reduction in greenhouse gas emissions.

Through a robust value creation plan, we helped CleanEnergyCo secure seed funding to build the foundation of the business, design a zero-based operating model, and nurture a dynamic company culture built around Agile working methods.

Mobilizing talent for a brand-new green energy venture

CleanEnergyCo's parent company had long aspired to lead the energy transition and meet the emerging needs of its customers. It had dabbled in renewables and hydrogen fuel, but its efforts in these areas had not achieved scale. The parent company's leaders envisioned an independent clean energy business that could operate with a new level of innovation and speed.

Working collaboratively with the parent company, our team helped define the core principles, structure, and financial plan for CleanEnergyCo. The parent company selected top talent to join the landing team that would jump-start the endeavor. Experts from Bain Academy provided training in the Agile methodologies that have proven to catalyze progress at start-ups and other new ventures.

With the talent thus mobilized, CleanEnergyCo set up its headquarters and began working with us on a comprehensive value creation plan.

Creating a blueprint for a scalable business

There were three main pillars to CleanEnergyCo's ambitions: renewable energy (e.g., solar, wind, and storage), green mobility (electric vehicle charging and fleet management), and clean hydrogen (production and distribution).

In each of these areas, we established clear pathways for CleanEnergyCo to become a market leader within the next decade. In each pillar, we helped the company design a detailed operating model with flexible structures, talent acquisition strategies, clear governance processes, and frameworks that encourage an innovation culture.

To accomplish all this, we applied our expertise across many different areas, including venture building from our Innovation & Business Building solution and access to the Venture Ecosystem, an ecosystem of investors and developers supporting next-generation solutions. We introduced CleanEnergyCo to a series of Bain ecosystem partners and advisers, including a start-up that uses data analytics to identify the optimal locations for electric vehicle (EV) charging hubs.

A core principle of building new ventures is not overdesigning for perfection; CleanEnergyCo is set up with a lean organizational structure, filling critical roles at the outset, with a defined process to ramp up gradually as the business grows.

A real solution takes the field

Upon its launch, CleanEnergyCo received a formal endorsement from the national government of the parent company. The company has activated the first EV chargers in its network, acquired stakes in multiple renewable energy projects and developers, and partnered with off-takers to produce and sell low-carbon hydrogen.

The strong growth potential and robust value creation plans have led the company's leaders to expand their long-term ambitions for CleanEnergyCo, multiplying their targets by a factor of five or more. Over the next 8 to 10 years, we expect the company to be one of the world's top producers of clean energy.

\* We take our clients' confidentiality seriously. While we've changed their names, the results are real.

## *Standardized*

### 1. Context

A leading oil and gas company had long aspired to play a stronger role in the energy transition. It had experimented with renewables and hydrogen-related initiatives, but these efforts had not achieved meaningful scale. The leadership envisioned creating an independent clean energy business capable of operating with greater innovation and speed.

### 2. Core Problem

The parent company wanted to build a clean energy solutions business but lacked a clear structure, operating model, and value creation plan for such an entity. Previous attempts in renewables and hydrogen remained fragmented and had not grown into a scalable business.

### 3. Situation at Start

At the beginning of the initiative, the parent company had no independent clean energy entity, only a strategic ambition to create one. Early explorations in renewables and hydrogen existed but were limited in impact. The company sought guidance on defining the principles, structure, and financial basis for a new venture and began selecting top internal talent to form a landing team.

### 4. Known Constraints

The new business needed to cover three main strategic areas: renewable energy such as solar, wind, and storage, green mobility such as electric vehicle charging and fleet management, and clean hydrogen including production and distribution

### 5. Objective

The parent company aimed to define, design, and launch a scalable clean energy venture and establish a clear blueprint for its long-term growth across renewable energy, green mobility, and clean hydrogen.

## Prompt Inputs and Model Outputs

### Baseline

#### *P01 Input:*

```
from openai import OpenAI

client = OpenAI(api_key="API_KEY_HERE")

P01 = """
You are given the following problem description.
Produce a plan for this case in four steps with 1-2 bullet points.
Do not provide final recommendations.
CASE DESCRIPTION:
A leading plastics producer operates seven global production sites.
Acetic acid is a key ingredient in many of its plastics manufacturing
processes and represents one of the company's largest variable cost
components.
The company has experienced consistently high acetic acid consumption
over several years and does not understand which factors within its
processes are driving this elevated usage.
At the beginning of the project, the company only knows that acetic
acid consumption has been high for an extended period and that the
underlying reasons remain unclear. Consumption levels vary, but no
explanation or systematic analysis has yet been carried out, and
internal teams suspect that process-related factors may be
contributing. A joint team of consultants, data scientists, and process
engineers is planned to address the issue.
The issue must be examined across all seven global sites, and any
understanding of improvement levers should preserve stable production
while addressing significant variable costs linked to acetic acid.
The company aims to identify the key factors that influence acetic acid
consumption and determine where opportunities may exist to reduce
usage.
"""

response = client.responses.create(
    model="gpt-5.1",
    input=P01,
```

```
        max_output_tokens=1200,  
        temperature=0.0  
    )  
  
print(response.output_text)  
print(response.model_dump_json(indent=2))
```

### *001 Output*

1. **\*\*Frame the problem and align scope\*\***
  - Clarify objectives, constraints (e.g., product quality, throughput, safety), and define common KPIs for acetic acid consumption across all seven sites.
  - Map major process steps where acetic acid is used and agree on hypotheses (e.g., operating conditions, equipment, raw material quality, control strategies) to be tested.
  
2. **\*\*Collect and prepare data across sites\*\***
  - Inventory and extract relevant data (process parameters, production volumes, product mix, maintenance logs, lab results) from all sites; standardize units, tags, and time stamps.
  - Clean, reconcile, and integrate datasets into a common analytical environment, ensuring data quality checks and documentation.
  
3. **\*\*Analyze drivers of consumption and quantify impact\*\***
  - Use statistical and advanced analytics (e.g., regression, clustering, time-series, ML) with process-engineering input to identify key variables and patterns linked to high acetic acid usage.
  - Compare sites and operating regimes to isolate best practices, structural differences, and controllable vs. non-controllable drivers, quantifying their impact on consumption.
  
4. **\*\*Design and test improvement levers\*\***
  - Co-develop with process engineers a set of potential levers (e.g., parameter setpoints, control logic changes, equipment or procedural adjustments) and prioritize them by impact and feasibility.
  - Plan controlled trials or simulations at selected sites to validate effects on acetic acid usage and production stability, and define requirements for broader rollout and monitoring.

```
{  
...
```

#### *P04 Input*

```
from openai import OpenAI
```

```
client = OpenAI(api_key="API_KEY_HERE")
```

```
P05 = """
```

You are given the following problem description.

Produce a plan for this case in four steps with 1-2 bullet points.

Do not provide final recommendations.

CASE DESCRIPTION:

A major APAC airline had gone through administration and was emerging from a period of severe industry and supplier disruptions caused by the Covid-19 pandemic. Travel usage patterns had shifted significantly due to reduced business travel, safety concerns, organizational downsizing, and decreased international flights. These changes affected demand, operations, and the supplier landscape.

The airline needed to achieve significant cost reductions and rebuild strong procurement capabilities after the pandemic and administration process.

Following administration, all contracts had been reset, requiring the airline to renegotiate multiple high-value, long-term agreements with a relatively new procurement team. The company faced urgent timelines driven by market conditions and nonnegotiable contract deadlines. Major procurement categories such as in-flight catering, ground handling, fuel, maintenance, repair and operations components, engine maintenance, and base maintenance were large cost drivers that required attention. The airline needed clarity on its spending base, contract terms, and procurement capability gaps.

Renegotiations had to occur under time pressure, influenced by market conditions, pandemic-related disruptions, and fixed contract deadlines. The procurement team was relatively new, and the airline operated across multiple high-value categories that had been reset during the administration process.

The airline aimed to reduce procurement costs and build best-in-class procurement capabilities to support stable operations after administration.

"""

```
response = client.responses.create(  
    model="gpt-5.1",  
    input=P05,  
    max_output_tokens=1200,  
    temperature=0.0  
)  
  
print(response.output_text)  
print(response.model_dump_json(indent=2))
```

### *004 Output*

1. Diagnose spend, contracts, and capabilities
  - Map end-to-end spend by category, supplier, and contract status; identify top value/risk contracts and deadline "cliff edges."
  - Assess current procurement skills, tools, governance, and data quality to pinpoint capability and process gaps post-administration.
  
2. Define category strategies and negotiation approach
  - Develop tailored strategies for key categories (e.g., catering, ground handling, fuel, MRO, engines, base maintenance) reflecting new demand patterns and market conditions.
  - Prioritize and sequence negotiations based on value at stake and timing constraints; define target outcomes, walk-away positions, and risk mitigations.
  
3. Execute rapid sourcing and renegotiation waves
  - Run structured sourcing events and negotiations for priority contracts, leveraging competitive tension, scenario modeling, and revised service/volume constructs.
  - Implement interim "bridge" arrangements where needed to avoid service disruption while securing longer-term, value-optimized agreements.
  
4. Build sustainable procurement capabilities and governance
  - Establish standardized category management, supplier management, and performance tracking processes supported by fit-for-purpose tools and data.

- Design and deliver targeted training, playbooks, and governance (e.g., approval thresholds, cross-functional deal teams) to embed best-in-class procurement practices.

{

...

### *P07 Input*

```
from openai import OpenAI
```

```
client = OpenAI(api_key="API_KEY_HERE")
```

```
P09 = """
```

```
You are given the following problem description.
```

```
Produce a plan for this case in four steps with 1-2 bullet points.
```

```
Do not provide final recommendations.
```

```
CASE DESCRIPTION:
```

```
A leading oil and gas company had long aspired to play a stronger role in the energy transition. It had experimented with renewables and hydrogen-related initiatives, but these efforts had not achieved meaningful scale. The leadership envisioned creating an independent clean energy business capable of operating with greater innovation and speed.
```

```
The parent company wanted to build a clean energy solutions business but lacked a clear structure, operating model, and value creation plan for such an entity. Previous attempts in renewables and hydrogen remained fragmented and had not grown into a scalable business.
```

```
At the beginning of the initiative, the parent company had no independent clean energy entity, only a strategic ambition to create one. Early explorations in renewables and hydrogen existed but were limited in impact. The company sought guidance on defining the principles, structure, and financial basis for a new venture and began selecting top internal talent to form a landing team.
```

```
The new business needed to cover three main strategic areas: renewable energy such as solar, wind, and storage, green mobility such as electric vehicle charging and fleet management, and clean hydrogen including production and distribution
```

```
The parent company aimed to define, design, and launch a scalable clean energy venture and establish a clear blueprint for its long-term growth across renewable energy, green mobility, and clean hydrogen.
```

```
"""
```

```
response = client.responses.create(  
    model="gpt-5.1",  
    input=P09,  
    max_output_tokens=1200,  
    temperature=0.0  
)  
  
print(response.output_text)  
print(response.model_dump_json(indent=2))
```

### *007 Output*

1. Diagnose current state and define ambition
  - Assess existing renewables, hydrogen, and mobility pilots, capabilities, and gaps; benchmark against leading clean energy players.
  - Clarify strategic ambition, risk appetite, and target role across renewables, green mobility, and clean hydrogen (e.g., developer, operator, integrator).
  
2. Design venture structure and operating model
  - Define legal/organizational setup, governance, decision rights, and interfaces with the parent company.
  - Design the operating model (business units, central platforms, talent model, partnerships, and innovation processes) to enable speed and scalability.
  
3. Shape portfolio and value creation blueprint
  - Prioritize focus segments in renewables, green mobility, and clean hydrogen based on market attractiveness and fit with capabilities.
  - Build a 5-10 year value creation plan including growth pathways, capital allocation, partnership strategy, and key performance metrics.
  
4. Plan launch and scale-up roadmap
  - Develop a phased launch plan (MVP scope, geographies, and pilots) and stand-up requirements (people, systems, processes).

- Define scale-up milestones, governance cadence, and mechanisms to periodically refine strategy and portfolio based on market feedback.  
{  
...

## Instruction Prompting

### *P02 Input*

```
from openai import OpenAI
```

```
client = OpenAI(api_key="API_KEY_HERE")
```

```
P02 = """
```

```
You are a strategy consultant. Develop a high-level, structured and actionable plan for how to analyze and solve the case described below.
```

```
Your output must:
```

- Present exactly 4 numbered main steps.
- For each step, briefly describe the key analytical action.
- Provide 1-2 short bullet points summarizing:
  - the most important data needed, and
  - the main intermediate insight or output.
- Focus only on essential elements and avoid unnecessary detail.
- Do not provide final recommendations. Focus only on the analytical approach.

```
CASE DESCRIPTION:
```

```
A leading plastics producer operates seven global production sites. Acetic acid is a key ingredient in many of its plastics manufacturing processes and represents one of the company's largest variable cost components.
```

```
The company has experienced consistently high acetic acid consumption over several years and does not understand which factors within its processes are driving this elevated usage.
```

```
At the beginning of the project, the company only knows that acetic acid consumption has been high for an extended period and that the underlying reasons remain unclear. Consumption levels vary, but no explanation or systematic analysis has yet been carried out, and internal teams suspect that process-related factors may be
```

contributing. A joint team of consultants, data scientists, and process engineers is planned to address the issue.

The issue must be examined across all seven global sites, and any understanding of improvement levers should preserve stable production while addressing significant variable costs linked to acetic acid.

The company aims to identify the key factors that influence acetic acid consumption and determine where opportunities may exist to reduce usage.

```
"""
```

```
response = client.responses.create(  
    model="gpt-5.1",  
    input=P02,  
    max_output_tokens=1200,  
    temperature=0.0  
)  
  
print(response.output_text)  
print(response.model_dump_json(indent=2))
```

## *002 Output*

### 1. Frame the problem and define analytical scope

Clarify what "high consumption" means, where it occurs, and how to measure improvement consistently across sites.

- Data needed: Historical acetic acid consumption and cost by site/plant/line, production volumes and product mix, existing process specs/targets, site-level constraints.

- Output: Clear problem statement, normalized KPIs (e.g., kg acetic acid per ton of product by product family/site), prioritized focus areas (sites/lines/products) for deeper analysis.

### 2. Build and clean an integrated data set across all sites

Consolidate operational, process, and quality data into a single analyzable structure, ensuring comparability across sites.

- Data needed: Time-stamped process parameters (temperatures, pressures, flow rates, residence times), batch/lot data, quality results, maintenance logs, shift/operator info, raw material specs.

- Output: Harmonized, validated data model with aligned tags/units, ready for statistical and machine-learning analysis; data quality assessment highlighting gaps and limitations.

### 3. Identify key drivers of acetic acid consumption

Use statistical and advanced analytics to uncover which factors most strongly explain variation in normalized consumption, controlling for production and product mix.

- Data needed: Integrated data set from step 2, including normalized consumption metrics and all candidate drivers (process settings, equipment status, ambient conditions, etc.).

- Output: Ranked list of significant drivers (e.g., specific process parameters, operating regimes, product types, equipment conditions), quantified impact sizes, and site-to-site comparison of driver relevance.

### 4. Translate drivers into improvement levers and test hypotheses

Convert analytical findings into concrete process hypotheses, then validate them through targeted deep dives and controlled tests while safeguarding production stability.

- Data needed: Detailed process flow diagrams, control logic/recipes, operating envelopes, safety/quality constraints, pilot or historical test data under different settings.

- Output: Shortlist of validated levers (e.g., parameter ranges, control strategies, equipment conditions) with estimated potential impact on acetic acid usage, and a prioritized set of experiments or trials for further validation.

{  
...

### *P05 Input*

```
from openai import OpenAI
```

```
client = OpenAI(api_key="API_KEY_HERE")
```

```
P06 = """
```

```
You are a strategy consultant. Develop a high-level, structured and actionable plan for how to analyze and solve the case described below.
```

Your output must:

- Present exactly 4 numbered main steps.
- For each step, briefly describe the key analytical action.
- Provide 1-2 short bullet points summarizing:
  - the most important data needed, and
  - the main intermediate insight or output.
- Focus only on essential elements and avoid unnecessary detail.
- Do not provide final recommendations. Focus only on the analytical approach.

#### CASE DESCRIPTION:

A major APAC airline had gone through administration and was emerging from a period of severe industry and supplier disruptions caused by the Covid-19 pandemic. Travel usage patterns had shifted significantly due to reduced business travel, safety concerns, organizational downsizing, and decreased international flights. These changes affected demand, operations, and the supplier landscape.

The airline needed to achieve significant cost reductions and rebuild strong procurement capabilities after the pandemic and administration process.

Following administration, all contracts had been reset, requiring the airline to renegotiate multiple high-value, long-term agreements with a relatively new procurement team. The company faced urgent timelines driven by market conditions and nonnegotiable contract deadlines. Major procurement categories such as in-flight catering, ground handling, fuel, maintenance, repair and operations components, engine maintenance, and base maintenance were large cost drivers that required attention. The airline needed clarity on its spending base, contract terms, and procurement capability gaps.

Renegotiations had to occur under time pressure, influenced by market conditions, pandemic-related disruptions, and fixed contract deadlines. The procurement team was relatively new, and the airline operated across multiple high-value categories that had been reset during the administration process.

The airline aimed to reduce procurement costs and build best-in-class procurement capabilities to support stable operations after administration.

""

```
response = client.responses.create(  
    model="gpt-5.1",  
    input=P06,  
    max_output_tokens=1200,  
    temperature=0.0  
)  
  
print(response.output_text)  
print(response.model_dump_json(indent=2))
```

### *005 Output*

#### 1. Diagnose spend, contracts, and demand shifts

Brief action: Build a clean, integrated view of post-administration spend, demand, and contracts to identify where to focus.

- Data needed: Historical and current spend by category/supplier; contract database (terms, volumes, expiry/renewal dates); updated demand forecasts by route/network and cabin.

- Output/insight: Prioritized category "heat map" showing value at stake, timing risk (deadlines), and disruption risk to guide sequencing of renegotiations.

#### 2. Assess category economics and supplier market conditions

Brief action: For each priority category, analyze cost drivers, market structure, and post-Covid supply-demand dynamics.

- Data needed: Benchmark prices and terms (industry data, peers, third-party indices); supplier financial health and capacity; operational performance and service levels.

- Output/insight: Category strategies (e.g., consolidate vs. diversify, insource vs. outsource, contract length, risk-sharing mechanisms) and target savings ranges by category.

#### 3. Design and execute accelerated negotiation playbooks

Brief action: Develop structured negotiation approaches and run parallel sourcing events aligned to deadlines and operational constraints.

- Data needed: Clean baseline of current total cost of ownership per supplier; should-cost models; switching costs and operational transition requirements; legal constraints from administration.

- Output/insight: Supplier-specific negotiation playbooks (value levers, walk-away points, concession ladders) and a time-phased negotiation and award plan across all major categories.

#### 4. Build sustainable procurement capabilities and governance

Brief action: Define the target operating model, tools, and skills needed for a best-in-class procurement function post-administration.

- Data needed: Current org structure, roles, and skill assessment; existing processes and policies; available systems/analytics tools and data quality.

- Output/insight: Procurement capability roadmap (org design, processes, tools, training) and governance model (category ownership, decision rights, performance KPIs) to sustain savings and manage future disruptions.

```
{  
...
```

#### *P08 Input*

```
from openai import OpenAI
```

```
client = OpenAI(api_key="API_KEY_HERE")
```

```
P10 = """
```

```
You are a strategy consultant. Develop a high-level, structured and actionable plan for how to analyze and solve the case described below.
```

```
Your output must:
```

- Present exactly 4 numbered main steps.
- For each step, briefly describe the key analytical action.
- Provide 1-2 short bullet points summarizing:
  - the most important data needed, and
  - the main intermediate insight or output.
- Focus only on essential elements and avoid unnecessary detail.
- Do not provide final recommendations. Focus only on the analytical approach.

```
CASE DESCRIPTION:
```

```
A leading oil and gas company had long aspired to play a stronger role in the energy transition. It had experimented with renewables and
```

hydrogen-related initiatives, but these efforts had not achieved meaningful scale. The leadership envisioned creating an independent clean energy business capable of operating with greater innovation and speed.

The parent company wanted to build a clean energy solutions business but lacked a clear structure, operating model, and value creation plan for such an entity. Previous attempts in renewables and hydrogen remained fragmented and had not grown into a scalable business.

At the beginning of the initiative, the parent company had no independent clean energy entity, only a strategic ambition to create one. Early explorations in renewables and hydrogen existed but were limited in impact. The company sought guidance on defining the principles, structure, and financial basis for a new venture and began selecting top internal talent to form a landing team.

The new business needed to cover three main strategic areas: renewable energy such as solar, wind, and storage, green mobility such as electric vehicle charging and fleet management, and clean hydrogen including production and distribution

The parent company aimed to define, design, and launch a scalable clean energy venture and establish a clear blueprint for its long-term growth across renewable energy, green mobility, and clean hydrogen."""

```
response = client.responses.create(  
    model="gpt-5.1",  
    input=P10,  
    max_output_tokens=1200,  
    temperature=0.0  
)  
  
print(response.output_text)  
print(response.model_dump_json(indent=2))
```

### *008 Output*

#### 1. Define ambition, scope, and value pools

Clarify what "success" looks like and where the new venture can win across renewables, green mobility, and clean hydrogen.

- Data needed: Market size/growth by segment and geography; competitor benchmarks; parent company assets/capabilities; regulatory/incentive landscape.
- Output/insight: Prioritized opportunity map and quantified ambition (e.g., target portfolio mix, revenue/EBIT targets, timing).

## 2. Design business architecture and portfolio choices

Translate ambition into a coherent business architecture, defining which businesses to build, partner, or exit and how they fit together.

- Data needed: Economics and risk profiles of candidate business models (e.g., develop-own-operate vs. originate-sell); customer segments and needs; synergy potential across segments.
- Output/insight: Target business portfolio (by segment and value chain position) and high-level business model designs for each pillar.

## 3. Define operating model and governance for the new entity

Specify how the clean energy venture will operate independently yet leverage the parent's strengths.

- Data needed: Current organizational structure, decision rights, and processes; talent inventory; existing shared services and technology platforms.
- Output/insight: Operating model blueprint (governance, org structure, decision rights, interfaces with parent, talent model, and key capabilities).

## 4. Build financial plan, roadmap, and launch plan

Develop the economic case, investment plan, and sequenced roadmap to stand up and scale the venture.

- Data needed: Capex/opex requirements by initiative; funding constraints; expected returns and risk scenarios; implementation capacity.
- Output/insight: Integrated financial model, prioritized initiative roadmap (waves of build-out), and launch plan with milestones and KPIs.

{  
...

## Decomposition Prompting

### *P03 Input*

```
from openai import OpenAI
```

```
client = OpenAI(api_key="API_KEY_HERE")
```

```
P03 = """
```

```
You are a strategy consultant. Develop a high-level, structured and actionable plan for how to analyze and solve the case described below.
```

```
First, think through the problem internally using the following steps:
```

1. Break the case into its fundamental analytical components, at a high level.
2. For each component, identify what must be understood or analyzed in order to address the overall case question.
3. For each component, determine what data or information would be most important to carry out that analysis.
4. Decide on a logical sequence in which to analyze these components, taking dependencies into account, and group them into exactly 4 main phases that will become the steps of your final plan.

```
Do NOT include these internal reasoning steps or this decomposition in the final answer.
```

```
Your output must:
```

- Present exactly 4 numbered main steps.
- For each step, briefly describe the key analytical action.
- Provide 1-2 short bullet points summarizing:
  - the most important data needed, and
  - the main intermediate insight or output.
- Focus only on essential elements and avoid unnecessary detail.
- Do not provide final recommendations. Focus only on the analytical approach.

```
CASE DESCRIPTION:
```

```
A leading plastics producer operates seven global production sites. Acetic acid is a key ingredient in many of its plastics manufacturing processes and represents one of the company's largest variable cost components.
```

```
The company has experienced consistently high acetic acid consumption over several years and does not understand which factors within its processes are driving this elevated usage.
```

At the beginning of the project, the company only knows that acetic acid consumption has been high for an extended period and that the underlying reasons remain unclear. Consumption levels vary, but no explanation or systematic analysis has yet been carried out, and internal teams suspect that process-related factors may be contributing. A joint team of consultants, data scientists, and process engineers is planned to address the issue.

The issue must be examined across all seven global sites, and any understanding of improvement levers should preserve stable production while addressing significant variable costs linked to acetic acid. The company aims to identify the key factors that influence acetic acid consumption and determine where opportunities may exist to reduce usage.

"""

```
response = client.responses.create(
    model="gpt-5.1",
    input=P03,
    max_output_tokens=1200,
    temperature=0.0
)

print(response.output_text)
print(response.model_dump_json(indent=2))
```

### *003 Output*

#### 1. Establish baseline performance and prioritize sites / products

Brief action: Quantify current acetic acid consumption patterns and identify where to focus deeper analysis.

- Data needed:

- Historical acetic acid consumption by site, line, product, batch; production volumes; product mix; yield; cost data.

- Intermediate output:

- Normalized benchmarks (e.g., kg acetic acid per ton of output) by site/product, with a prioritized list of "high-usage" sites/lines for detailed investigation.

#### 2. Map processes and define hypothesized drivers of consumption

Brief action: Build a structured view of how acetic acid is used and what process factors could influence its consumption.

- Data needed:

- Process flow diagrams, P&IDs, operating manuals; standard recipes and setpoints; maintenance logs; operator practices; quality specs and scrap/rework data.

- Intermediate output:

- Driver tree linking acetic acid usage to process steps and potential drivers (e.g., temperature, pressure, residence time, impurities, rework), with a shortlist of priority hypotheses per site/line.

### 3. Conduct quantitative driver analysis and identify improvement levers

Brief action: Use statistical and engineering analysis to quantify which factors most strongly explain consumption variability.

- Data needed:

- Time-series process data (DCS/SCADA): key operating parameters, alarms, downtime; batch records; raw material quality; environmental conditions; site-level constraints.

- Intermediate output:

- Ranked list of statistically and technically validated drivers of acetic acid consumption, with estimated impact ranges and indicative improvement levers (e.g., parameter optimization, recipe adjustments, loss reduction).

### 4. Validate findings, test changes, and build implementation roadmap

Brief action: Confirm causality through trials and translate insights into a practical, risk-managed change plan.

- Data needed:

- Pilot / trial results from parameter changes; before-after consumption and quality data; operational risk assessments; implementation cost and feasibility inputs from sites.

- Intermediate output:

- Validated set of levers with quantified savings potential, site-by-site implementation roadmap, and monitoring KPIs to track acetic acid usage and process stability over time.

{

## *P06 Input*

```
from openai import OpenAI
```

```
client = OpenAI(api_key="API_KEY_HERE")
```

```
P06 = """
```

```
You Are a strategy consultant. Develop a high-level, structured and  
actionable plan for how to analyze and solve the case described below.
```

```
First, think through the problem internally using the following steps:
```

1. Break the case into its fundamental analytical components, at a high level.
2. For each component, identify what must be understood or analyzed in order to address the overall case question.
3. For each component, determine what data or information would be most important to carry out that analysis.
4. Decide on a logical sequence in which to analyze these components, taking dependencies into account, and group them into exactly 4 main phases that will become the steps of your final plan.

```
Do NOT include these internal reasoning steps or this decomposition in  
the final answer.
```

```
Your output must:
```

- Present exactly 4 numbered main steps.
- For each step, briefly describe the key analytical action.
- Provide 1-2 short bullet points summarizing:
  - the most important data needed, and
  - the main intermediate insight or output.
- Focus only on essential elements and avoid unnecessary detail.
- Do not provide final recommendations. Focus only on the analytical approach.

```
CASE DESCRIPTION:
```

```
A major APAC airline had gone through administration and was emerging  
from a period of severe industry and supplier disruptions caused by the  
Covid-19 pandemic. Travel usage patterns had shifted significantly due  
to reduced business travel, safety concerns, organizational downsizing,
```

and decreased international flights. These changes affected demand, operations, and the supplier landscape.

The airline needed to achieve significant cost reductions and rebuild strong procurement capabilities after the pandemic and administration process.

Following administration, all contracts had been reset, requiring the airline to renegotiate multiple high-value, long-term agreements with a relatively new procurement team. The company faced urgent timelines driven by market conditions and nonnegotiable contract deadlines. Major procurement categories such as in-flight catering, ground handling, fuel, maintenance, repair and operations components, engine maintenance, and base maintenance were large cost drivers that required attention. The airline needed clarity on its spending base, contract terms, and procurement capability gaps.

Renegotiations had to occur under time pressure, influenced by market conditions, pandemic-related disruptions, and fixed contract deadlines. The procurement team was relatively new, and the airline operated across multiple high-value categories that had been reset during the administration process.

The airline aimed to reduce procurement costs and build best-in-class procurement capabilities to support stable operations after administration.

"""

```
response = client.responses.create(  
    model="gpt-5.1",  
    input=P06,  
    max_output_tokens=1200,  
    temperature=0.0  
)  
  
print(response.output_text)  
print(response.model_dump_json(indent=2))
```

### *006 Output*

1. Diagnose post-administration spend, demand, and contract baseline

Brief action: Build a clean, fact-based view of what the airline buys, from whom, under what terms, and how demand has structurally changed post-Covid.

- Key data needed:

- Historical and current spend by category/supplier/route; volumes (flights, passengers, sectors) by cabin and route; current and legacy contract terms (prices, SLAs, indexation, volumes, exit clauses).

- Main intermediate output:

- Consolidated "as-is" baseline: category spend cube, demand and service-level requirements by route/network, contract heatmap (value, risk, urgency, flexibility) to identify priority categories and contracts.

## 2. Assess category economics, market conditions, and value levers

Brief action: For each priority category (e.g., fuel, MRO, catering, ground handling), analyze cost drivers, supplier markets, and savings levers under current APAC and global conditions.

- Key data needed:

- Category cost breakdowns (e.g., fuel vs fees vs labor), supplier market structure and capacity, benchmark prices/terms, Covid/post-Covid demand and capacity trends, regulatory/airport constraints.

- Main intermediate output:

- Category strategies and value hypotheses: quantified savings ranges, key negotiation levers, risk/continuity constraints, and preferred sourcing models (e.g., single vs multi-supplier, regional vs local).

## 3. Design and sequence the renegotiation and sourcing program

Brief action: Translate category strategies into a time-bound, risk-managed negotiation and sourcing roadmap aligned with contract deadlines and operational needs.

- Key data needed:

- Contract expiry/renewal dates, operational criticality by contract, supplier performance data, internal resource capacity, legal/administration constraints, implementation lead times.

- Main intermediate output:

- Integrated sourcing and negotiation plan: prioritized wave schedule, negotiation playbooks by category, target savings and risk

thresholds, and decision frameworks for trade-offs (cost vs resilience vs service).

#### 4. Build and embed best-in-class procurement capabilities

Brief action: Identify capability gaps in the new procurement team and design the operating model, tools, and governance needed to sustain performance post-program.

- Key data needed:

- Current org structure and roles, skill and experience assessment, existing processes and policies, systems/tools landscape (e.g., P2P, contract management, analytics), performance KPIs.

- Main intermediate output:

- Procurement capability blueprint: target operating model, prioritized capability build plan (skills, processes, tools), governance and KPI framework, and change-management plan to institutionalize new ways of working.

{

...

#### *P09 Input*

```
from openai import OpenAI
```

```
client = OpenAI(api_key="API_KEY_HERE")
```

```
P09 = """
```

```
You Are a strategy consultant. Develop a high-level, structured and actionable plan for how to analyze and solve the case described below.
```

```
First, think through the problem internally using the following steps:
```

```
1. Break the case into its fundamental analytical components, at a high level.
```

```
2. For each component, identify what must be understood or analyzed in order to address the overall case question.
```

```
3. For each component, determine what data or information would be most important to carry out that analysis.
```

```
4. Decide on a logical sequence in which to analyze these components, taking dependencies into account, and group them into exactly 4 main phases that will become the steps of your final plan.
```

Do NOT include these internal reasoning steps or this decomposition in the final answer.

Your output must:

- Present exactly 4 numbered main steps.
- For each step, briefly describe the key analytical action.
- Provide 1-2 short bullet points summarizing:
  - the most important data needed, and
  - the main intermediate insight or output.
- Focus only on essential elements and avoid unnecessary detail.
- Do not provide final recommendations. Focus only on the analytical approach.

CASE DESCRIPTION:

A leading oil and gas company had long aspired to play a stronger role in the energy transition. It had experimented with renewables and hydrogen-related initiatives, but these efforts had not achieved meaningful scale. The leadership envisioned creating an independent clean energy business capable of operating with greater innovation and speed.

The parent company wanted to build a clean energy solutions business but lacked a clear structure, operating model, and value creation plan for such an entity. Previous attempts in renewables and hydrogen remained fragmented and had not grown into a scalable business.

At the beginning of the initiative, the parent company had no independent clean energy entity, only a strategic ambition to create one. Early explorations in renewables and hydrogen existed but were limited in impact. The company sought guidance on defining the principles, structure, and financial basis for a new venture and began selecting top internal talent to form a landing team.

The new business needed to cover three main strategic areas: renewable energy such as solar, wind, and storage, green mobility such as electric vehicle charging and fleet management, and clean hydrogen including production and distribution

The parent company aimed to define, design, and launch a scalable clean energy venture and establish a clear blueprint for its long-term growth across renewable energy, green mobility, and clean hydrogen.

""

```

response = client.responses.create(
    model="gpt-5.1",
    input=P09,
    max_output_tokens=1200,
    temperature=0.0
)

print(response.output_text)
print(response.model_dump_json(indent=2))

```

### *009 Output*

#### 1. Clarify strategic ambition, scope, and value-creation thesis

Briefly define what "success" looks like and where/how the new entity should win.

- Data needed: Corporate strategy and decarbonization targets; existing renewables/hydrogen pilots (business cases, performance); competitor benchmarks and investor expectations; regulatory/market outlook for renewables, mobility, hydrogen.

- Output/insight: Clear strategic north star (mission, target markets, risk/return profile), prioritized role across renewables, green mobility, and hydrogen, and a preliminary value-creation thesis (where profit pools are, how the entity will differentiate).

#### 2. Define portfolio strategy and business model choices

Determine which businesses to build, partner, or exit, and how they will make money.

- Data needed: Market size/growth and economics by segment (solar, wind, storage, EV charging, fleet, hydrogen value chain); customer needs and willingness to pay; competitive landscape; internal capabilities and assets (sites, grid connections, customer base, trading, engineering).

- Output/insight: Target portfolio (which segments, geographies, customer types), chosen business models (developer/owner-operator/platform/service provider), and high-level financial archetypes (returns, capital intensity, scaling path) for each pillar.

### 3. Design the entity structure, operating model, and governance

Translate strategy into an organizational blueprint that enables speed and innovation while leveraging the parent.

- Data needed: Current org structure, decision rights, and processes; talent inventory and skills; existing JV/partnership structures; parent's risk, compliance, and capital allocation frameworks.

- Output/insight: Proposed legal/organizational setup (degree of independence), operating model (functions, roles, decision rights, interfaces with parent), governance model (board, investment committees), and critical capabilities/talent requirements.

### 4. Build the financial plan, roadmap, and implementation blueprint

Quantify the plan and define how to launch and scale over time.

- Data needed: Capex/opex benchmarks by asset type; financing options and cost of capital; pipeline of potential projects; internal funding constraints; implementation risks and dependencies.

- Output/insight: Integrated 5-10 year financial plan (P&L, capex, portfolio build-out), phased roadmap (waves of businesses/markets, milestones), and implementation plan (landing team mandate, quick wins, partnership plan, and key risk mitigations).

{

...