



**Leiden University**

**ICT in Business and the Public Sector**

**Strategic Foresight in the Age of AI:  
Exploring the Role of Generative AI in  
Management Consulting**

Name: Ricardo Ruud Johannes Schreuder BSc

Student ID: s2526301

Date: 30/10/2025

Supervisor 1: Dr. M. Hilbert

Supervisor 2: Dr. C.J. Stettina MSc

Company Supervisor: Joost Elffers

**MASTER'S THESIS**

Leiden Institute of Advanced Computer Science (LIACS)

Leiden University

Einsteinweg 55

2333 CC Leiden

The Netherlands

## **Abstract**

The rapid emergence of Generative Artificial Intelligence (AI) has begun to transform knowledge-intensive sectors such as management consulting. Yet, its precise role in supporting strategic planning activities remains underexplored. This thesis examines where generative AI can and cannot support strategic planning activities in management consulting, identifying both benefits and challenges.

An exploratory sequential mixed methods design was employed. Six expert interviews with consultants from SAP and KPMG first identified the strategic planning activities and narrowed the six-step strategic planning model to two critical phases, Environmental Analysis and Strategy Evaluation, by asking interviewees to rank where generative AI would be most suitable. The subsequent quantitative phase, based on 128 survey responses from KPMG consultants, was used to generalize these findings. The survey quantified, for each activity, current satisfaction and future importance. Then the activity-level gap scores were computed to locate where generative AI support appears most promising versus limited. Finally, a practical demonstration applied structured prompting strategies to a SWOT analysis to translate the empirical insights into applied consulting practice.

The interviews converged on Environmental Analysis and Strategy Evaluation as the most amenable phases for generative AI support; the survey confirmed a marked gap between current satisfaction and future importance, indicating potential not yet realized in practice. Three benefits stand out: (1) mature use cases in structured analytical frameworks, such as SWOT, PESTLE and Porter's Five Forces, (2) promising but still developing uses in market, competitor and risk analyses and technological trend forecasting and (3) a "sparring partner" role where generative AI enhances creativity and accelerates idea generation without displacing human judgment. The most challenging activities are feasibility assessment, scenario creation and scenario evaluation given their reliance on contextual understanding and human judgment. Lastly, an underdeveloped level of prompting literacy among consultants was identified, which constrains the value in practice.

By combining empirical evidence with an applied demonstration, this thesis shows that generative AI already adds tangible value to selected strategic planning activities while remaining constrained by contextual and interpretive limits. The study contributes to both theory and practice by clarifying where generative AI's integration into consulting workflows is most effective and by establishing a foundation for future research on its role in strategic management.

## **Acknowledgments**

Writing this thesis has been a challenging yet deeply rewarding experience that allowed me to grow both academically and personally. Throughout the process, I have had the privilege of receiving guidance, support and encouragement from several people, to whom I would like to express my sincere gratitude.

First and foremost, I would like to sincerely thank my first supervisor, Dr. Marc Hilbert, for his clear guidance, constructive feedback and structured approach throughout this journey. Even during times when progress seemed uncertain, your structured approach and steady support helped bring clarity and momentum back into the process.

I would also like to thank my second supervisor Dr. Christoph Stettina, whose thorough feedback and reflective discussions during this research helped me connect the empirical findings with broader insights. Your perspective added significant depth to the final work.

A special word of thanks goes to the SAP and KPMG consultants who participated in the interviews and survey. Your openness and insights formed the foundation of this study and made it possible to translate abstract concepts into practical relevance.

I am deeply grateful to my family and friends for their unwavering support and understanding. Lastly, I am thankful for the assistance of ChatGPT [60], which was solely used for proofreading my thesis, specifically for grammar, spelling errors and textual coherence.

# Contents

<b>1</b>	<b>Introduction</b>	<b>8</b>
1.1	Problem statement . . . . .	8
1.2	Research gap . . . . .	9
1.3	Relevance . . . . .	9
1.4	Thesis outline . . . . .	10
<b>2</b>	<b>Literature review</b>	<b>11</b>
2.1	Generative AI . . . . .	11
2.1.1	Large Language Models . . . . .	12
2.1.2	History of LLMs . . . . .	12
2.1.3	ChatGPT . . . . .	14
2.1.4	Prompting . . . . .	15
2.1.4.1	Background and historical development . . . . .	16
2.1.4.2	Prompting strategies . . . . .	17
2.1.4.3	Prompting best practices . . . . .	26
2.1.4.4	Prompting as a skill . . . . .	27
2.2	Strategic planning . . . . .	29
2.2.1	Definition and purpose . . . . .	29
2.2.2	Practitioners of strategic planning . . . . .	30
2.2.3	Management consulting . . . . .	31
2.2.4	Process of strategic planning . . . . .	32
2.2.5	Strategic planning activities . . . . .	35
2.2.5.1	SWOT analysis . . . . .	35
2.2.5.2	PESTLE analysis . . . . .	36
2.2.5.3	Stakeholder analysis . . . . .	36
2.2.5.4	Porter’s Five Forces . . . . .	36
2.2.5.5	Operational tasks . . . . .	37
<b>3</b>	<b>Methodology</b>	<b>39</b>
3.1	Research design . . . . .	39
3.2	Qualitative phase . . . . .	40
3.2.1	Data collection . . . . .	41
3.2.2	Sampling technique . . . . .	41
3.2.3	Data analysis . . . . .	42
3.2.4	Validation & reliability . . . . .	42

3.3	Quantitative phase . . . . .	43
3.3.1	Data collection . . . . .	44
3.3.2	Sampling technique . . . . .	45
3.3.3	Data analysis . . . . .	45
3.3.3.1	Gap analysis . . . . .	46
3.3.3.2	Software . . . . .	46
3.3.4	Validation & reliability . . . . .	46
3.4	Practical demonstration . . . . .	47
<b>4</b>	<b>Results</b>	<b>48</b>
4.1	Qualitative results . . . . .	48
4.1.1	Participants' demographics . . . . .	48
4.1.2	Content analysis . . . . .	49
4.1.3	Ranking of strategic planning steps . . . . .	50
4.2	Survey results . . . . .	53
4.2.1	Demographics . . . . .	54
4.2.2	Attitudes toward generative AI . . . . .	56
4.2.3	Current state of prompting and enablement . . . . .	58
4.2.4	Gap Analysis of current satisfaction & future importance . . . . .	60
4.2.5	In-depth gap analysis based on demographics . . . . .	65
<b>5</b>	<b>Discussion</b>	<b>68</b>
5.1	Generative AI's role in strategic planning . . . . .	68
5.1.1	Benefits . . . . .	68
5.1.2	Challenges . . . . .	72
5.2	Prompting as a critical but underdeveloped competence . . . . .	75
5.3	Limitations . . . . .	76
5.4	Future research . . . . .	77
<b>6</b>	<b>Conclusion</b>	<b>79</b>
	<b>References</b>	<b>79</b>
	<b>Appendices</b>	<b>87</b>
	<b>Appendix A Interview guide</b>	<b>87</b>
	<b>Appendix B Survey instrument</b>	<b>91</b>
	B.1 Open-ended question answers . . . . .	97
	<b>Appendix C Practical demonstration</b>	<b>103</b>
	C.1 Prompt engineering . . . . .	105

<b>Appendix D</b>	<b>Output of LLMs</b>	<b>107</b>
D.1	Stage 1 – Baseline SWOT . . . . .	107
D.2	Stage 2 – Refined prompt . . . . .	109
D.3	Stage 3 – Reasoning overlay (Chain-of-Thought) . . . . .	112
D.4	Stage 4 – Few-shot micro exemplars . . . . .	116
D.5	Stage 5 – Grounded variant (Simulated RAG) . . . . .	120
D.6	Stage 6 – Targeted refinement & finalization . . . . .	125
<b>Appendix E</b>	<b>Figures</b>	<b>130</b>

# List of Figures

2.1	Zero-shot iteration 1, a naïve prompt results in generic and shallow insights, resembling a quick online search. . . . .	18
2.2	Zero-shot iteration 2: refinement through role, style and tone leads to quantified, trade-off oriented insights that go beyond generic commentary. . . . .	18
2.3	Few-shot iteration 1, one basic exemplar improves structure over zero-shot but still lacks segmentation, quantification and actionable mitigations. . . . .	19
2.4	Few-shot iteration 2, the refined exemplar teaches the full consulting template, now adding segmentation, a risk–mitigation table, quantified ROI and OKRs for a decision-ready deliverable. . . . .	20
2.5	Chain-of-Thought iteration 1, explicit trade-offs surface the reasoning, but the argument is still compact and largely qualitative. . . . .	22
2.6	Chain-of-Thought iteration 2, an explicit argument map makes the reasoning auditable, where premises, inferences and implications lead to a defensible decision. . . . .	23
2.7	Retrieval-Augmented Generation iteration 1, external retrieval grounds claims but lacks per-claim traceability. . . . .	24
2.8	RAG iteration 2, adds claim-level traceability, light quantification and enforceable decision conditions vs. iteration 1. . . . .	25
2.9	The six-step strategic planning process . . . . .	33
4.1	Number of respondents indicating 'Not applicable / Not familiar with' per strategic activity. . . . .	54
4.2	Age distribution . . . . .	54
4.3	Department distribution . . . . .	54
4.4	Experience level distribution . . . . .	55
4.5	Function types distribution . . . . .	55
4.6	Distribution of AI usage . . . . .	56
4.7	Strategic planning frequency . . . . .	56
4.8	Attitude scores by Experience . . . . .	57
4.9	Attitude scores by department . . . . .	57
4.10	Attitude scores by age . . . . .	57
4.11	Attitude scores by function . . . . .	57
4.12	Share of respondent have received internal or external training on the use of Generative AI. . . . .	58
4.13	Usefulness AI training . . . . .	59

4.14	Need for prompting strategies . . . . .	59
4.15	Consultant’s support needs for Generative AI use . . . . .	60
4.16	Current prompting practices among consultants . . . . .	60
4.17	Distribution of current satisfaction and future importance scores among all consultants. . . . .	61
4.18	Gap analysis illustrating the gap scores, including satisfaction and importance for all strategic planning activities. . . . .	63
4.19	Gap analysis illustrating the gap scores, including satisfaction and importance for all strategic planning activities by each corresponding step. . . . .	65
4.20	Heatmap of gap, satisfaction, and importance scores for each strategic planning activity, broken down by age group. Values represent mean scores per age. Darker shades indicate more extreme values within each scale. . . . .	67
D.1	Stage 1 – Zero-shot SWOT with broad generic items and overlap. . . . .	108
D.2	Stage 2 – Role and strict schema make items concise and consistent, part 1 . . .	110
D.3	Stage 2 – Role and strict schema make items concise and consistent, part 2 . . .	111
D.4	Stage 3 – Examples guide tone and raise specificity for this domain, part 1 . . .	113
D.5	Stage 3 – Examples guide tone and raise specificity for this domain, part 2 . . .	114
D.6	Stage 3 – Examples guide tone and raise specificity for this domain, part 3 . . .	115
D.7	Stage 4 – Stepwise prompting clarifies logic and reduces quadrant leakage, part 1	117
D.8	Stage 4 – Stepwise prompting clarifies logic and reduces quadrant leakage, part 2	118
D.9	Stage 4 – Stepwise prompting clarifies logic and reduces quadrant leakage, part 3	119
D.10	Stage 5 – Grounded in source snippets so claims are evidence based, part 1 . . .	121
D.11	Stage 5 – Grounded in source snippets so claims are evidence based, part 2 . . .	122
D.12	Stage 5 – Grounded in source snippets so claims are evidence based, part 3 . . .	123
D.13	Stage 5 – Grounded in source snippets so claims are evidence based, part 4 . . .	124
D.14	Stage 6 – Quality checking the output of the LLM and refining it, part 1 . . . .	126
D.15	Stage 6 – Quality checking the output of the LLM and refining it, part 2 . . . .	127
D.16	Stage 6 – Quality checking the output of the LLM and refining it, part 3 . . . .	128
D.17	Stage 6 – Quality checking the output of the LLM and refining it, part 4 . . . .	129
E.1	Heatmap of gap, satisfaction, and importance scores for each strategic planning activity, broken down by functional role. Values represent mean scores per function. Darker shades indicate more extreme values within each scale. . . . .	130
E.2	Heatmap of gap, satisfaction, and importance scores for each strategic planning activity, broken down by experience group. Values represent mean scores per experience group. Darker shades indicate more extreme values within each scale.	131
E.3	Heatmap of gap, satisfaction, and importance scores for each strategic planning activity, broken down by department. Values represent mean scores per department. Darker shades indicate more extreme values within each scale. . . .	132

# List of Tables

2.1	From zero-shot to RAG: comparative overview of prompting strategies in consulting . . . . .	26
4.1	Descriptive characteristics and interview duration of the interviewees . . . . .	48
4.2	Strategic planning activities identified from expert interviews, grouped by planning step. . . . .	49
4.3	Top-3 rankings of strategic planning steps per interviewee. . . . .	51
4.4	Converted scores using Modified Borda Count. . . . .	52
4.5	Activities identified for Environmental Analysis and Strategy Evaluation in the interview phase. . . . .	52
4.6	Gap analysis of generative AI usefulness across strategic planning activities. <i>M_Sat</i> and <i>M_Imp</i> indicate mean satisfaction and importance; <i>Gap</i> is the mean of respondent-level paired difference (Imp - Sat); <i>SD_Sat</i> , <i>SD_Imp</i> , and <i>SD_Gap</i> report standard deviations; <i>n</i> is the number of responses. . . . .	62
4.7	Top 3 and Bottom 3 activities by average standard deviation across satisfaction, importance, and gap . . . . .	63
B.1	Respondents' perspectives on integrating Generative AI in strategic planning activities. . . . .	97
B.2	Respondents' examples of situations where Generative AI supported strategic planning activities and what worked well. . . . .	99
B.3	Respondents' reported limitations and risks when using Generative AI for strategic-planning activities. . . . .	100
B.4	Consultants' suggestions for improving the effective use of Generative AI in strategic planning. . . . .	101
B.5	Consultants' self-reported prompting practices when using Generative AI tools. . . . .	101
C.1	Overview of prompting stages, benefits and limitations . . . . .	106

# Chapter 1 Introduction

## 1.1 Problem statement

This section outlines the central issue addressed by this thesis. It begins by briefly describing the current problem and why it is relevant. Next, the significance of this issue is highlighted from both academic and business perspectives, establishing a clear foundation for the subsequent research focus.

Generative Artificial Intelligence (AI) has rapidly emerged as a transformative technology, providing the opportunity to reshape business operations [63]. It automates formerly manual knowledge-intensive tasks such as report generation, customer engagement and data analysis, which increases operational efficiency and scalability [63]. In addition, generative AI gathers and integrates vast datasets to provide structured insights for strategic decision-making. These shifts are equally significant for management consulting, where the synthesis of complex information and the delivery of strategic recommendations form the core of practice, ultimately allowing consultants to focus on strategic activities of higher value [44]. Essentially, generative AI transforms business operations by shifting human effort away from repetitive information processing toward judgment, creativity and strategic choice. Large Language Models (LLMs), including ChatGPT, Claude, Gemini and others are able to understand and generate human-like text and much more, creating opportunities for innovation in multiple sectors. This research will focus on ChatGPT as the main LLM, as the researcher has upgraded access to all its advanced features, is peer-reviewed in various studies and has widespread adoption [85, 44]. Recent studies indicate that generative AI can significantly enhance efficiency and insight generation in fields like finance, healthcare, retail and customer service [63]. Despite these advances, its potential to specifically support strategic planning activities in management consulting remains underexplored.

Strategic planning is a structured process that organizations use to define their long-term goals and determine the best ways to achieve them [13]. It uses tools and frameworks, such as SWOT analysis, scenario planning, competitors and stakeholder analysis, the analysis of internal success factors and cognitive mapping to inform decision making [7, 82, 83]. The aforementioned tools and frameworks can be called activities of strategic planning. Traditionally, strategic planning was predominantly undertaken by top executives and dedicated planning teams. However, recent research underscores the value of involving multiple layers of the organization, ranging from junior staff to middle managers, in the planning process [55, 84, 34]. This broader inclusion reduces the gap between high-level decisions and operational realities.

Management consultants are integral to this process, leveraging specialized knowledge to guide organizations in formulating and executing strategic plans. They typically help with

implementing strategic planning tools, ensure alignment with organizational objectives and provide an outsider's perspective to challenge internal assumptions [82, 47]. In a climate of accelerating technological change management consultants are not only tasked with expert business advice, but also to deal with navigating clients through digital transformations and emerging technologies. Therefore, this field may be well-suited for the capabilities of generative AI.

This focus on technological augmentation is not entirely new. The potential for AI in strategic planning has already long been recognized. Orsini recognized the potential of AI for strategic planning as early as 1986, stating that "Artificial Intelligence is the tool that Strategic Planning needs to fulfill its expectations," way before the widespread adoption of AI in late 2022 [62]. The current generative AI models may be able to live up to these expectations.

## 1.2 Research gap

Although strategic planning, management consulting and its activities are well-documented, there is notable absence of research on how generative AI could be utilized to enhance these activities specifically within the technological advisory sector. Preliminary investigations using key terms like '*management consulting*', '*strategic planning*', '*generative AI*', '*generative AI in management consulting*' and '*ChatGPT and consulting*' revealed no substantial relevant studies. Recent work has begun to analyze AI's potential in broader consulting contexts, highlighting benefits such as increased efficiency and improved decision-making, alongside challenges like data privacy and ethical considerations [69, 44]. However, these studies do not delve into the unique potential benefits and challenges of where generative AI can support the strategic planning activities. Moreover, as mentioned by Fayaz et al. it is underexplored how generative AI affects strategic planning within management consulting [57].

Addressing this gap is crucial for both academic and industry stakeholders. Understanding where generative AI can be deployed to support strategic planning activities in management consulting may offer insights for consultants, clients and researchers. For instance, LLMs could automate initial data analysis, identify emerging trends in real time or generate first-pass recommendations for scenario planning, enabling consultants to focus on higher-value strategic decisions.

This leads to the following research question:

**RQ.** *What are the benefits and challenges of using Generative AI to support strategic planning activities in management consulting?*

## 1.3 Relevance

This research is relevant both academically and practically. From an academic perspective, it addresses a clear gap in the literature concerning the application of generative AI, specifically

Large Language Models (LLMs), to strategic planning within the management consulting domain. While previous studies have explored AI in broader consulting contexts, few have focused on its potential to support specific planning activities. By applying a mixed methods approach, including survey-based analysis and practical demonstrations, this thesis offers a novel contribution that bridges theoretical insights with real-world applicability and lays the groundwork for future studies in this emerging field.

From a practical standpoint, this research is conducted at KPMG, a global consulting firm operating at the intersection of strategy and technology. The findings offer actionable insights into where LLMs can be used to support strategic planning activities in day-to-day consulting practice. By identifying both benefits and challenges, the study helps consultants make informed decisions about the strategic fit and potential of generative AI within their workflows, providing the foundation for improved efficiency and the quality of strategic recommendations.

## **1.4 Thesis outline**

The outline of this thesis is as follows: Chapter 2 presents a literature review that covers the foundations of generative AI, prompting techniques, strategic planning activities and the role of management consultants.

Chapter 3 outlines the methodology. It details the exploratory sequential mixed methods design, including the qualitative interviews, survey construction and a practical demonstration using ChatGPT. Chapter 4 presents the results from both phases, including, content analysis, activity rankings and gap analysis.

Chapter 5 discusses the findings and its limitations and places them in context. Chapter 6 concludes the thesis with a conclusion and future research directions. The practical demonstration is shown in Appendix C, as it is not directly related to the data and research question, though provides added value for practical insights.

# Chapter 2 Literature review

This chapter reviews existing literature to establish a theoretical foundation for understanding the role of generative AI in strategic planning within management consulting. It introduces and defines key concepts relevant to this study including generative AI, Large Language Models (LLMs), ChatGPT and prompting. This technological foundation is followed by a detailed examination of strategic planning, its underlying process, the role of management consultants and its activities that characterize strategic planning work. This broad theoretical analysis serves as the foundation for the rest of the research.

## 2.1 Generative AI

Generative AI has rapidly emerged as a transformative technology that is changing the way how businesses operate [63]. Generative AI is capable of producing original content across a range of modalities, including text, images, audio and video. As a subset of Artificial Intelligence, generative AI systems use advanced machine learning techniques, primarily deep learning, to model complex patterns in data and generate new outputs that resemble human-like material. These systems are trained on vast amounts of data and operate autonomously once trained, enabling rapid and scalable content creation. While specific applications of generative AI vary across industries, their core function remains the same: to generate content that mimics human output.

Within the broader family of deep learning architectures, transformer-based models have become the dominant force behind text generation [80]. Although other architectures like Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs) contributed to early advances in visual and audio generation, they are less applicable to the language-focused tasks relevant to this study [86]. The focus here is on models that generate and manipulate natural language, a capability that is at the core of strategic communication and analysis [13].

The rise of generative AI tools has come in parallel with the rapid expansion of the global AI market, which experienced a growth rate of 31% in 2023 alone [24]. According to recent projections by *Statista*, the global AI industry is expected to exceed 800 billion euros by 2030, with generative AI accounting for approximately 350 billion of that total. These figures highlight not only economic momentum but also its already existing and growing prominence.

Current tools span a wide spectrum of capabilities. For example, DALL-E, DeepAI and Canva enable users to create visuals based on textual input. Similarly, in audio synthesis, platforms such as ElevenLabs can replicate human voices with remarkable accuracy and generate original soundtracks. In video generation, AI models like Sora 2 and Runway produce high-quality videos based on textual prompts, sometimes achieving results nearly indistinguishable

from real-life footage. However, these multimodal applications are primarily used in creative domains such as marketing, entertainment and design. In contrast, text-based generative models are more directly relevant in the decision-oriented analytics context, such as strategic planning in management consulting [44]. These tools can support tasks like analyzing trends, summarizing data, simulating business scenarios and performing risk analysis.

Among all generative AI applications, Large Language Models (LLMs) stand out as the most widely adopted. Models such as ChatGPT, Copilot and Gemini excel in understanding and generating human-like text. Interest in these LLMs has risen since the launch of ChatGPT by OpenAI in late 2022, offering new opportunities for automation, decision-making and strategic innovation across various sectors [63].

### **2.1.1 Large Language Models**

LLMs are a prominent category of generative AI systems capable of understanding and generating human-like text. They can perform a wide range of tasks, including but not limited to, content generation, translation, conversational interaction, reasoning and even programming support [1]. LLMs excel in interpreting context, identifying patterns and producing fluent and coherent responses, making them valuable in domains that require knowledge synthesis and textual articulation. LLMs are already widely adopted across sectors such as finance, healthcare, retail and customer service, where they are used to enhance efficiency and user experience [63]. Their impact is especially notable in customer service, where they automate repetitive interactions, improve response accuracy and enable personalized communication at scale [63]. These capabilities mirror the growing need for tools that can support complex, information-intensive activities such as those found in strategic planning and consulting.

### **2.1.2 History of LLMs**

The development of LLMs is grounded in decades of progress in AI and computational linguistics. The conceptual origins can be traced back to the 1940s when Warren McCulloch and Walter Pitts introduced the idea of an Artificial Neural Network (ANN) [66]. In subsequent decades, the first language models were developed, primarily based on rule-based approaches. At the time, Natural Language Processing (NLP) tasks were still relatively simple and mostly limited to binary classification problems, such as identifying whether a word belonged to a specific category [66].

The 1990s marked a significant advancement with the development of statistical language models. These models were capable of processing vast amounts of text data, allowing for greater accuracy compared to their rule-based predecessors. Statistical models enabled early NLP applications, including text prediction, speech recognition and spam detection. However, they remained restricted in their ability to capture semantic relationships and contextual nuances. A key innovation occurred in the mid-2000s with the introduction of word embeddings, a technique that allowed language models to understand semantic relationships through numerical

representations [37]. Although word embeddings improved language understanding, they still lacked contextual awareness and often struggled with homonyms and homophones. These are straightforward for humans to interpret, but challenging for computers because the same word, such as 'match', can have multiple meanings depending on the context.

The 2010s saw a major shift in LLM development with the emergence of neural language models. Notably, Google's Neural Machine Translation (GNMT) model was among the first to integrate neural networks for NLP, significantly outperforming traditional rule-based and statistical models in terms of accuracy and fluency. It enabled the translation of complex texts while preserving linguistic structures. The true turning point, however, came in 2017 with the introduction of the Transformer architecture model by Vaswani et al. [50, 80]. The Transformer model utilized a self-attention mechanism, allowing it to analyze the relationships between words across an entire sentence, rather than processing words sequentially. This approach enabled faster training, better contextual understanding and greater scalability [80]. Unlike previous models, that struggled with longer text dependencies, Transformers were capable of capturing long-term relationships between words due to their ability to process information in parallel. Processing information in parallel means executing multiple operations simultaneously. This increased computational efficiency gave rise to the development of LLMs with billions of parameters.

Following the success of Transformers, researchers began developing pre-trained LLMs, which can be fine-tuned for specific applications. The first major pre-trained LLM were Generative Pre-Trained Transformer (GPT) released by OpenAI and Bidirectional Encoder Representations from Transformers (BERT) developed by Google [50, 66]. OpenAI played a key role in advancing LLM technology through its GPT series. OpenAI's GPT models became widely known for their accessibility and performance across diverse use cases. The evolution of the GPT series reflects rapid scaling and architectural refinement:

- **GPT-1** (2018) was trained on 117 million parameters, which are learned numerical values predicting what word is next after a sequence of words. This process is constantly optimized during training [50].
- **GPT-2** (2019) significantly expanded upon its predecessor, trained on 1.5 billion parameters, allowing for more coherent and contextually aware text generation and enabling more complex text generation.
- **GPT-3** (2020) marked another significant leap, trained on 175 billion parameters. It introduced zero-shot and few-shot learning capabilities, allowing the model to perform tasks without needing explicit retraining for each one [66, 10].
- **GPT-3.5** (2022) is a fine-tuned version of GPT-3, incorporating and combining multiple AI advancements. These were deep learning, unsupervised pre-training, instruction fine-tuning, multi-task learning, in-context learning and Reinforcement Learning from Human Feedback (RLHF) [85]. These technologies together enabled GPT-3.5 to generate highly

coherent, context-aware and human-like responses, making it the first widely adopted consumer-facing LLM. This version also marked the introduction of ChatGPT, which was at that time based on GPT-3.5.

- **GPT-4** (2023) the last of this series was trained on an enormous set of 1.77 billion parameters, having introduced multimodal capabilities, such as interpreting text, audio and images and enabling web browsing [1].

### 2.1.3 ChatGPT

ChatGPT is the conversational variant of OpenAI's GPT series, which evolved from GPT-1 through GPT-4 and has since extended into additional modal lines [66, 85]. In late 2022, OpenAI publicly released GPT-3.5, marking the first time an advanced LLM was made accessible through an intuitive chat interface. ChatGPT gained global attention due to its ease of use and human-like conversational abilities, leading to 100 million monthly active users within just two months, making it one of the fastest-adopted consumer products in history [85]. This success catalyzed interest in generative AI's potential for both individual productivity and organizational transformation.

ChatGPT is capable of having human-like multi-turn conversations, answering a vast range of queries. However, early concerns arose regarding the accuracy and reliability of its responses, as ChatGPT occasionally generated factually incorrect or misleading information [86]. These limitations highlighted the need for critical oversight when applying LLMs to decision-oriented domains, such as consulting and strategic planning. The latest ChatGPT version of the GPT series to date is GPT-4, which has web-browsing capabilities, allowing real-time information retrieval. Additionally, it introduces enhanced multimodal capabilities, enabling users to input text, images, PDFs and audio files [85]. Having these multimodal capabilities introduces new tasks it can perform, such as image captioning, paper summarizing and chart reasoning [85]. GPT-4 is able to outperform other LLMs in a variety of NLP tasks, in English as well as other languages, further demonstrating its capabilities [1].

Following GPT-4, OpenAI launched a new family of models, known as the "o" series. The o stands for "omnimodel" and is designed for more seamless and efficient multimodal interactions, o3-pro is its latest version. These models are optimized for chain-of-thought (CoT) reasoning, enabling more step-by-step analysis in complex tasks. Compared to GPT-4, the "o" models are particularly favored for reasoning-heavy tasks but are less optimal for language-focused applications due to longer response times caused by CoT reasoning. The currently deployed models within ChatGPT are the GPT series and "o" series; GPT-4o, GPT-4.5, o3, o4-mini and o4-mini-high, each offering distinct strengths in terms of speed, reasoning or multimodal interaction. As OpenAI has continued innovating with flagship models, so have others.

The rapid adoption of ChatGPT sparked industry-wide competition, prompting companies like Google and Meta to develop their own publicly available LLMs. Google has since introduced the Gemini (formerly Bard) and Gemma models, while Meta has launched the LLaMA model

family, both designed to compete within the evolving AI landscape [53, 78]. Google's Gemini represents Google Deepmind's latest advancement in AI reasoning, Retrieval-Augmented Generation (RAG) and multimodal capabilities. One of the features the Gemini model stands out is its integration of real-time web search [78]. Furthermore, the Gemini family includes advanced models like LearnLM, which is specifically designed based on the pedagogical literature to optimize learning via a LLM. This makes it highly effective for learning and tutoring use cases, offering a more advanced form of AI-based instruction than simply prompting the LLM to behave as a tutor [78]. Meta's LLaMa is a pre-trained and instruction-fine-tuned model, demonstrating advanced reasoning, instruction following and code generation [53]. Moreover, Meta's LLaMA, in contrast, positions itself as the most open-source aligned model, though its licensing restricts commercial use and imposes geographical and branding limitations that conflict with the Open Source Definition [53]

While Gemini, LLaMa and ChatGPT each bring unique features to the evolving LLM ecosystem, ChatGPT remains the most widely used and widely studied. Its integration into sectors such as customer service, finance and healthcare is already established [63]. In the context of management consulting, its capacity to rapidly synthesize information, generate insights and perform a variety of analyses underscores its practical relevance. Though LLMs are not a replacement for expert judgment, recent research demonstrates their augmentative potential. In one study comparing the performance of a human analyst and ChatGPT on a strategic business task, the human analyst outperformed the model, but improved their own performance by 12.5% when using ChatGPT as a support tool [41]. This finding reinforces the notion that LLMs like ChatGPT, when applied carefully, can support and enhance the effectiveness of strategic planning activities.

#### **2.1.4 Prompting**

LLMs produce meaningful and context-aware responses based on the input they receive, referred to as a prompt [73]. While early prompting relied solely on text, recent multimodal LLMs can process combinations of text, images, audio and documents, expanding the scope and complexity of user interactions [73]. Prompting thus refers to the formulation of these inputs in a way that steers the model's behaviour toward desired outcomes. The wording, structure and intent embedded within a prompt directly influence the model's relevance and accuracy [49]. As such, the practice of prompt engineering, reflects the need for deliberate, iterative refinement of inputs to optimize LLM performance. Rather than passively querying the model, prompt engineering involves designing, testing and adjusting prompts based on feedback and iteration. This practice has proven essential for high-stakes, knowledge-intensive domains such as strategic planning, where generic outputs or misinterpretations can undermine decision quality [14]. Prompting is now widely recognized as a key digital competence. It blends technical skills (e.g., understanding model behaviour), cognitive abilities (e.g., problem structuring) and creativity (e.g., framing open-ended or role-based instructions) to guide AI output effectively [46, 49, 73]. In strategic

contexts, prompting proficiency directly affects the utility and reliability of AI-generated insights. As a result, the ability to prompt well is no longer optional; it is increasingly viewed as a critical skill for professionals interacting with generative AI. This section introduces prompting as a foundational mechanism for working with LLMs. The remainder of the chapter explores its historical evolution, key prompting strategies, practical best practices and its formal recognition as a professional skill.

#### **2.1.4.1 Background and historical development**

The concept of prompting has developed in parallel with advancements in LLMs. Early approaches to NLP were dominated by supervised learning paradigms that required fine-tuning a model for each specific task. However, as also outlined in 2.1.2, the introduction of the transformer architecture enabled the creation of more scalable and generalizable LLMs. Initial transformer-based models, such as BERT and GPT introduced the possibility of using prefix inputs or natural language templates to steer model outputs, although within relatively fixed boundaries [23, 65].

A transformative shift occurred with the release of GPT-3 in 2020, which popularized the pre-train, prompt and predict paradigm, widely considered as a turning point in human-AI (LLMs) interaction design [49]. In this new paradigm, the model’s behaviour is steered not through additional supervised training, but by reformulating tasks as natural language prompts. This eliminates the need for explicit fine-tuning of internal parameter engineering and allows the model to generate task-relevant outputs directly based on its pre-trained knowledge. Prompting thus became the primary method for guiding the behaviour of models without changing their internal parameters, enabling their application across a wide range of tasks, including strategic planning.

While early uses of prompting by Fan et al., experimented with input phrasing, the concept of “prompt engineering” was only formally articulated as a methodological practice following GPT-3’s release, particularly by Reynolds & McDonell [68, 10, 28]. This marked the start of systematic studies into prompting strategies, such as zero-shot, few-shot and Chain-of-Thought (CoT) prompting, which formed the foundational backbone of modern prompt design [73, 10].

These foundational techniques have since evolved into a wide range of prompting strategies. For instance, Sahoo et al. identify 41 distinct prompt engineering techniques, Liu et al. report 26 and Schulhoff et al. list 58, underscoring the field’s rapid expansion [73, 70, 49]. Each of these works emphasizes the importance of selecting an appropriate prompting strategy depending on the task and domain-specific context.

Consequently, prompt engineering has become a vital skill and a critical digital competence. Writing effective prompts now requires a combination of technical precision, domain knowledge and cognitive effort [46]. In knowledge-intensive activities, such as strategic planning, where contextual understanding and goal alignment are crucial, the ability to engineer prompts meaningfully influences the value of AI-generated outputs.

#### 2.1.4.2 Prompting strategies

As aforementioned the output of a LLM is significantly influenced by how its prompt is formulated. This has led to the development of various prompting techniques recently, each with its own uniqueness and purpose. Within strategic planning, a domain characterized by its knowledge-intensive sector it's crucial to design a prompt in order to retrieve the best possible outcome. Drawing from academic literature, four prompting strategies are chosen: zero-shot, few-shot, Chain-of-Thought (CoT) and Retrieval-Augmented Generation (RAG) [70, 73]. To ensure consistency across the examples, all prompting strategies are applied to the same illustrative case:

*A European technology consulting firm is considering launching a “Managed Generative AI Advisory Service” as a new offering for enterprise clients. The aim of the service is to help organizations adopt and govern generative AI solutions in a compliant and scalable manner.*

The examples are presented in a hybrid structure. First, iteration within the zero-shot strategy demonstrates how refinement improves quality. Thereafter, more advanced strategies build upon the previous outputs: few-shots addresses the limits of zero-shot, CoT adds structured reasoning and Retrieval-Augmented Generation (RAG) grounds results in external evidence. This approach shows both intra-strategy iteration and intra-strategy escalation (i.e., building upon previous strategies), reflecting how prompting develops in practice.

Zero-shot is a simple prompting method, it refers to presenting the model with only an instruction or question, without any examples. This minimalist approach relies entirely on the model's pre-trained knowledge, depending purely on carefully crafted prompts to guide the output [65, 70]. While simple to apply zero-shot prompts require careful phrasing to ensure clarity and avoid ambiguous results. Prompt structure can be optimized by assigning roles, specifying tone, structure guidance or embedding stylistic cues [70]. This method may be well-suited for preliminary tasks, such as performing a basic PESTLE analysis, idea generation and early-stage brainstorming, which are exploratory scenarios providing initial guidance.

In Figure 2.1, the zero-shot prompt produces a generic output with limited structure and no quantification. On reviewing this, the prompt was refined to address the identified shortcomings: the role of a senior consultant was introduced to raise the professional standard, the style was framed in consulting language to align with client expectations and explicit case context was added to reduce generic phrasing. The refined prompt, Figure 2.2, produces a more nuanced feasibility analysis, including indicative figures such as team size and break-even horizon and presents trade-offs in an advisory tone. This demonstrates the iterative process of diagnosing weaknesses in an initial output and modifying the prompt accordingly, a critical digital competence for consultants using generative AI.

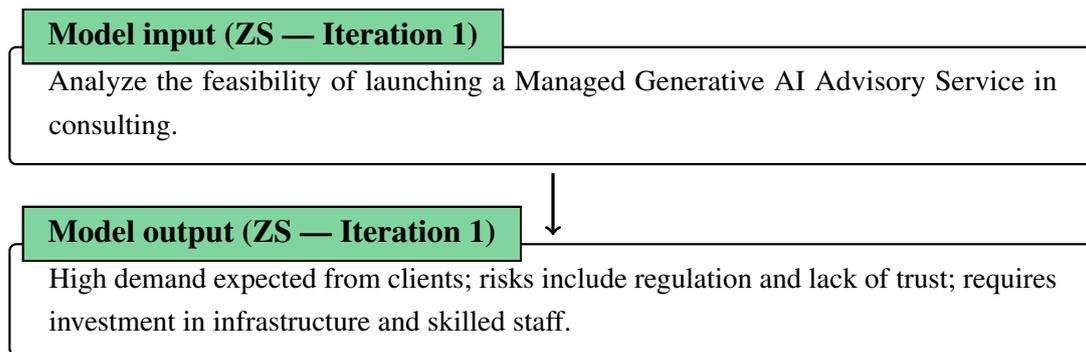


Figure 2.1: Zero-shot iteration 1, a naïve prompt results in generic and shallow insights, resembling a quick online search.

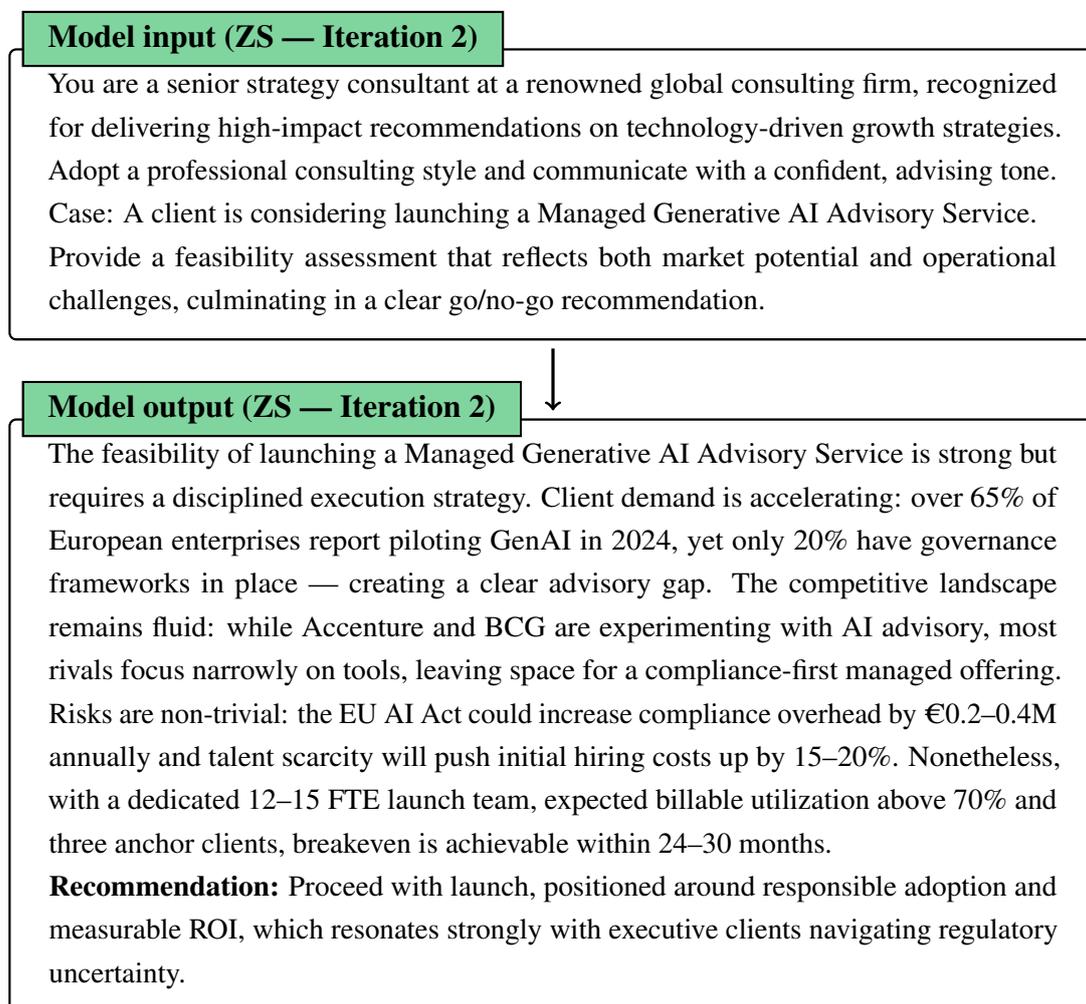


Figure 2.2: Zero-shot iteration 2: refinement through role, style and tone leads to quantified, trade-off oriented insights that go beyond generic commentary.

In contrast to zero-shot prompting, few-shot prompting relies less on prompt precision, as it provides a few input-output examples to give the model more context, before providing the actual query [10]. This improves accuracy by guiding the model’s response but requires more

tokens than zero-shot prompting [70]. This approach enhances performance in activities that benefit from structure or patterns, as this model is already presented with likewise tasks on how to solve them. This method may be well-suited for generating a competitor analysis that follows a predefined format, where the consultant provides a few prior examples to ensure consistency across different industries or clients.

In Figure 2.3, the model receives one input-output example before being given the actual task. This helps the LLM replicate the analysis style and provides clearer structure than in the zero-shot case, but it does not clearly outperform the refined zero-shot iteration. This demonstrates an important point: a well-crafted zero-shot prompt can already produce consulting-grade insights and a minimal few-shot example may not add much depth on its own. The output remains fairly high-level, lacking segmentation, assumptions and risk mitigations.

Figure 2.4 shows the second iteration of few-shot prompting, where a refined, more comprehensive exemplar is used to calibrate both style and analytical depth. This refinement directly addresses the shortcomings of the first iteration: the model now introduces demand segmentation, articulates assumption and provides a structured risk. Here the value of few-shot prompting becomes clear: the model produces a standardized consulting deliverable aligned with the exemplars. By embedding best-practice formats into the prompt, consultants can ensure outputs consistently follow established strategic assessment standards. Unlike the zero-shot case, the few-shot approach ensures consistency of format and depth, which is critical in consulting practice.

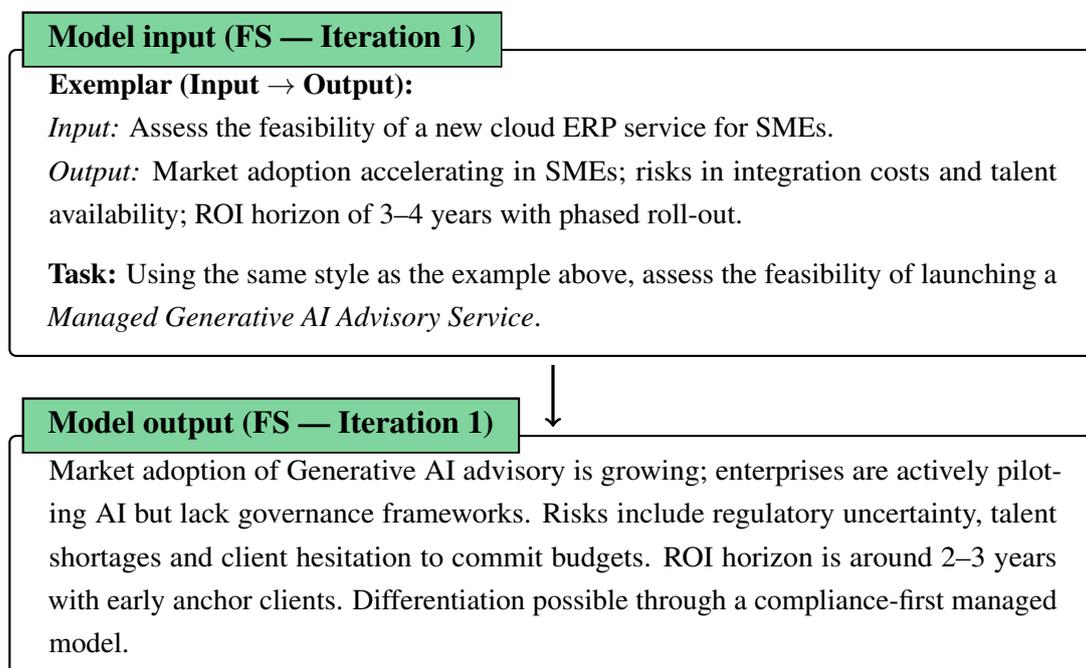


Figure 2.3: Few-shot iteration 1, one basic exemplar improves structure over zero-shot but still lacks segmentation, quantification and actionable mitigations.

### Model input (FS — Iteration 2)

#### Exemplar A (Input → Output)

*Input:* Assess a managed **AI customer support assistant** for software companies.

*Output (format to learn):*

**Executive view:** Viable for software vendors with high ticket volumes seeking faster replies.

**Demand segmentation:** - SaaS with many “how-to” questions: strong fit. - On-premise software: moderate fit, fewer tickets.

**Risk–mitigation (1 row example):** Wrong answers ⇒ restrict to approved knowledge base and add supervisor review.

**ROI horizon:** 12–18 months with three mid-market clients.

**Key OKR (Q1):** Cut average first response time by 40% on two pilot accounts.

**Differentiation:** Quick integration with helpdesk tools and clear quality dashboards.

#### Task (apply the same structure learned from the exemplars):

Assess the feasibility of launching a Managed Generative AI Advisory Service. Return the same sections as in the exemplars: (i) Executive view, (ii) Demand segmentation, (iii) Risk–mitigation table, (iv) ROI horizon, (v) Key OKRs (first 12 months), (vi) Differentiation.



### Model output (FS — Iteration 2)

**Executive view:** Launching a Managed GenAI Advisory Service is feasible and timely. The market is expanding rapidly and demand for compliance-ready AI guidance is underserved.

**Demand segmentation:** - *Financial services:* High demand for AI governance frameworks. - *Healthcare:* Strong need for compliant workflows and documentation. - *Manufacturing:* Moderate adoption, tied to predictive maintenance and quality control pilots.

#### Risk–mitigation table (excerpt):

Risk	Likelihood	Mitigation
Compliance over-head	High	Establish governance processes upfront
Talent scarcity	Medium	Targeted recruitment, partnerships with AI vendors
Client skepticism	Medium	Proof-of-value pilots and ROI case studies

**ROI horizon:** Breakeven achievable in 24–30 months with  $\geq 3$  anchor clients and 70%+ utilization.

**Key OKRs (first 12 months):** Q1: secure 2 lighthouse clients; Q2: utilization  $\geq 65\%$ ; Q3: compliance framework operational; Q4: NPS  $\geq 50$ .

**Differentiation:** Position as the trusted partner for responsible GenAI adoption, leveraging compliance-first delivery, managed operations and measurable ROI.

Figure 2.4: Few-shot iteration 2, the refined exemplar teaches the full consulting template, now adding segmentation, a risk–mitigation table, quantified ROI and OKRs for a decision-ready deliverable.

For more complex reasoning tasks, CoT prompting has been introduced, as LLMs often struggle with complex reasoning [81]. CoT encourages LLMs to break down problems into intermediate steps. This improves logical reasoning and improves the results in areas such as mathematics and structured decision-making [70]. This prompting method also presents an opportunity to understand how the model came to its output [81]. This is beneficial for identifying hallucination or logical errors, as can often happen with LLMs [86]. This method may be well-suited for assessing strategy feasibility, where consultants must break down the process into sequential steps, such as evaluating legal barriers, assessing competitive intensity, estimating cost structures and forecasting ROI, to ensure a transparent and justifiable recommendation.

As can be seen in Figure 2.5, CoT prompts the LLM to break down the complex task into distinct steps, explicitly outlining the required structure for the output. Unlike few-shot, it doesn't provide a full example but rather instructs the model on how to approach the problem, leading to a more structured and reasoned response. However, the output remains descriptive, counterarguments are not considered and assumptions are left implicit.

After review, the prompt was refined to require a brief argument map with explicit assumptions, step-by-step reasoning and short responses to likely objections. The revised version in Figure 2.6 delivers a clear, checkable line of reasoning that supports a defensible go/no-go decision. This illustrates iterative refinement: diagnosing weaknesses in the first pass and adjusting the prompt to produce more transparent, decision-ready logic an essential digital competence.

### Model input (CoT — Iteration 1)

You are an senior strategy consultant. Step by step, reason through the feasibility of launching a Managed Generative AI Advisory Service, building on the few-shot assessment:

- (1) relate demand to regulatory risk,
- (2) relate delivery capacity to talent constraints,
- (3) relate cost structure to ROI horizon,
- (4) conclude with a recommendation grounded in these trade-offs.

### Model output (CoT — Iteration 1)

**Demand vs. regulation.** Demand is strongest where governance is mandatory (finance, healthcare). Regulatory exposure is a risk, but a compliance-first posture converts it into a differentiator.

**Capacity vs. talent.** Talent scarcity is binding in the short term; narrowing initial scope to two sectors raises utilization and protects delivery quality.

**Cost vs. ROI.** Upfront costs are unavoidable, yet a managed model turns advisory into recurring value, stabilizing returns after the ramp.

**Recommendation.** Proceed with a focused launch in two regulated sectors; stage hiring to signed anchors; make governance the centrepiece of the value proposition.

Figure 2.5: Chain-of-Thought iteration 1, explicit trade-offs surface the reasoning, but the argument is still compact and largely qualitative.

### Model input (CoT — Iteration 2)

Act as a strategy partner. Using chain-of-thought, present your reasoning as a concise argument map for the feasibility of a Managed Generative AI Advisory Service, building on the few-shot assessment. Use exactly these sections in your final output (no scratch work): (1) Framing question, (2) Explicit assumptions (3–5 bullets), (3) Reasoning steps (premise → inference → implication; 5–7 lines), (4) Counterarguments & responses (2–3), (5) Decision & rationale (≤4 sentences). Avoid heavy numbers; focus on logic, trade-offs and causality.

### Model output (CoT — Iteration 2)

**Framing question.** Is a compliance-first *Managed Generative AI Advisory Service* feasible and defensible as a new offering?

**Explicit assumptions.** • Enterprises are piloting AI faster than they can build governance.

- Regulated sectors value auditability and policy templates.
- Early delivery capacity is limited; focus is required.
- Clients prefer low-risk starts (pilots) with clear success criteria.

**Reasoning steps (premise → inference → implication).** 1) Many pilots lack governance → buyers seek trusted guidance → an advisory service centred on governance is valued.

2) Regulation raises perceived risk → providers who operationalize compliance reduce buyer anxiety → compliance-first positioning differentiates.

3) Scarce talent limits breadth → concentrate on two sectors with similar needs → higher utilization and consistent quality.

4) One-off projects reset learning → managed engagement preserves context → recurring value and stickier client relationships.

5) Executives need proof before scale → pilot-led entry with clear success metrics → faster internal buy-in and referenceability.

6) Competing firms market tooling → advisory that ties process, policy and onboarding fills a gap → defensible space against tool-only vendors.

**Counterarguments & responses.** • “*Regulation could delay adoption.*” Response: pilots scoped to policy-compliant use cases create a safe on-ramp.

• “*Talent is too scarce.*” Response: narrow sector focus + partner bench offsets early constraints.

• “*Clients may build in-house.*” Response: governance frameworks and operating rhythms transfer, but managed oversight remains valuable.

**Decision & rationale.** Proceed with a compliance-first managed advisory focused on two regulated sectors. The argument map shows demand rooted in governance gaps, a clear differentiation against tool-centric rivals and an operating model that converts advisory into recurring value. Scope, sequence and governance gates control downside while credibility accumulates through pilot-led wins.

Figure 2.6: Chain-of-Thought iteration 2, an explicit argument map makes the reasoning auditable, where premises, inferences and implications lead to a defensible decision.

While the previous techniques utilize existing pre-trained knowledge, RAG enhances LLM capabilities by retrieving and incorporating external data sources before generating an answer. This retrieved information is then integrated into the prompt. This prompting technique is used for knowledge-intensive tasks, as the LLM now does not primarily have to rely on its training data [73]. For example, the LLM may first browse the web before responding, enabling it to provide real-time, fact-based answers [70]. As a result, RAG is well-suited for strategic planning activities such as market trend analysis, where relying solely on training data could result in outdated or incomplete outputs [70]. Moreover, by grounding responses in external facts, RAG reduces the likelihood of hallucinations, instances where the model generates plausible-sounding but incorrect information, thereby improving factual accuracy and decision reliability.

The last example, as shown in Figure 2.7, includes an explicit instruction to 'retrieve' external information before generating the response. This indicates that the LLM is expected to fetch and integrate real-time or specific external knowledge into its answer, differentiating it by its reliance on up-to-date, external data sources. However, this iteration does not yet provide an in-depth answer on the feasibility of launching a Managed Generative AI Advisory Service.

Iteration 2 of RAG (Figure 2.8) provides an in-depth, decision-ready analysis rather than a high-level summary. It introduces claim–source tags ([S1]–[S4]) and a consolidated source list at the end of the output, so each material statement can be audited. The output also marks, per statement, which source underpins the claim and where available includes direct links to the underlying documents, enabling immediate fact-checking. This structure materially improves traceability, reduces hallucination risk and clarifies feasibility by tying evidence to scope, risks and Go/No-Go conditions.

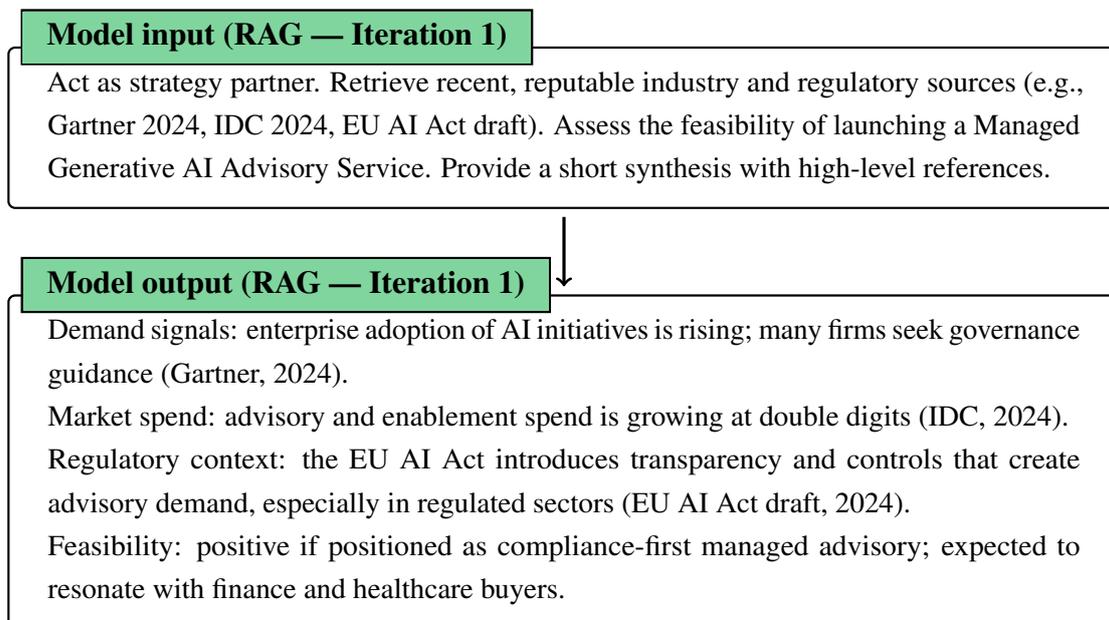


Figure 2.7: Retrieval-Augmented Generation iteration 1, external retrieval grounds claims but lacks per-claim traceability.

### Model input (RAG — Iteration 2)

Act as a strategy partner. **Retrieve** (date  $\geq 2024$ ) from reputable sources (e.g., Gartner, IDC, European Commission/EU AI Act text, ISO/IEC standards). Produce a **Decision Brief** for launching a Managed Generative AI Advisory Service (EU focus). **Requirements:** A) Demand & spend signals (qualitative or ranges), B) Buyer ICPs & triggers (who buys, why, when), C) Regulatory implications (obligations shaping scope), D) Risks & mitigations, E) Go/No-Go with conditions. For every material claim, attach a bracketed key [S1]–[S4] and list sources at the end with titles.



### Model output (RAG — Iteration 2)

#### Decision Brief: Managed GenAI Advisory (EU, 2026 launch)

**A) Demand & spend signals.** Enterprise GenAI programs are moving from pilots to governed scale; AI services spend growing at double-digit rates across EMEA [S1][S2].

**B) Buyer ICPs & triggers.** Primary ICPs: regulated enterprises (FSI, healthcare, public) with multi-model estates. Triggers: board-level AI strategy mandates; upcoming audits; vendor consolidation; post-pilot production risk reviews [S1][S2].

**C) Regulatory implications (scope shaping).** EU AI Act introduces obligations around transparency, risk management, data governance and post-market monitoring, with phased enforcement; ISO/IEC AI management system standards formalize control frameworks [S3][S4].

**D) Risks & mitigations.** Scope creep  $\rightarrow$  fixed discovery + control catalog [S3]. Model/provider lock-in  $\rightarrow$  multi-model reference architecture [S1]. Value capture risk  $\rightarrow$  KPI tree (use-case ROI, time-to-value) and stage gates [S2].

**E) Go/No-Go (conditions).** **Go** if: (1) verticalized control library (FSI, healthcare) is ready; (2) partner ecosystem covers audit/assurance; (3) pricing aligns to outcomes (pilot, scale). **No-Go** if: client lacks data governance owner or refuses minimal documentation [S3].

#### Claim–source traceability

- “Double-digit AI services growth”  $\rightarrow$  [S2]
- “Governed scale replacing pilots”  $\rightarrow$  [S1]
- “Transparency/risk management obligations”  $\rightarrow$  [S3]
- “Management system approach to AI controls”  $\rightarrow$  [S4]

#### Sources:

S1) Gartner (2024): Enterprise GenAI adoption and governance trends.

S2) IDC (2024): AI services/advisory spend growth outlook (EMEA).

S3) European Commission / EU AI Act (2024): risk-based obligations and timelines.

S4) ISO/IEC (2024): AI management system / governance standards guidance.

Figure 2.8: RAG iteration 2, adds claim-level traceability, light quantification and enforceable decision conditions vs. iteration 1.

Across all four strategies, iteration 1 reveals concrete shortcomings (generic scope in zero-shot, limited depth/consistency in few-shot, implicit assumptions in CoT and non-auditable evidence in RAG). Iteration 2 then addresses these gaps in a targeted way: role/tone/context raise quality in zero-shot; exemplars standardize client-ready structure in few-shot; argument maps make reasoning transparent in CoT; and per-claim citations with conflict checks make RAG auditable. Together, these iterations show that prompt engineering is an iterative craft. Deliberate refinements produce considerable gains in structure, reasoning transparency and evidentiary traceability. This establishes prompt engineering as a critical digital competence for consultants using generative AI.

Table 2.1 provides a clear overview of the prompting strategies: zero-shot, few-shot, CoT and RAG. By displaying its typical output, limitations and strategic value in consulting.

Prompting Strategy	Typical Output	Limitation	Strategic Value in Consulting
<b>Zero-shot</b>	Generic, exploratory insights (e.g., opportunities/risks list), improve by introducing: role, tone and guidance	Shallow, lacks sector nuance	Quick brainstorming, starting point
<b>Few-shot</b>	Structured outputs following exemplars (e.g., competitor matrix)	Requires high-quality examples; risk of copying bias	Template-consistent deliverables; repeatable analyses
<b>CoT</b>	Stepwise reasoning with explicit assumptions	May overcomplicate simple tasks	Transparency, traceability, scenario evaluation
<b>RAG</b>	Evidence-based outputs citing external sources	Requires curated knowledge base	Traceability, credibility, grounding in real-time data

Table 2.1: From zero-shot to RAG: comparative overview of prompting strategies in consulting

These prompting strategies demonstrate that prompt formulation directly influences a LLM effectiveness, making prompt engineering a critical factor in AI-driven communication and decision-making [70]. While these unique prompting strategies have been explained and illustrated, they can also be combined to retrieve a more optimal output. For instance, CoT can be integrated with few-shot prompting, leading to the development of few-shot CoT. Therefore, it is important to carefully consider a tailored prompting strategy that suits the desired goals.

### 2.1.4.3 Prompting best practices

Effective prompt engineering is crucial for maximizing the capabilities of LLMs. A well-designed prompt increases the relevance, factual accuracy and contextual awareness of the model's output. Drawing from the literature, several best practices have been found that guide prompt formulation to optimize performance and reduce the risk of misleading or hallucinated responses.

Foremost among these is the imperative to craft prompts with clarity and specificity. Ambiguous or vague prompts often result in incoherent or overly general outputs. Precision in phrasing is crucial for ensuring the model's complete comprehension of the desired output [70, 73]. A particularly strong technique involves assigning a distinct role or persona to the LLM (e.g., "Act as a management consultant" or "You are an expert in the area of strategic planning"). Such strategic guidance refines the model's response style, content and narrows its focus to a specific domain, thereby aligning closer to the user's intent [16, 70]. Furthermore, when using examples within prompts it's crucial that they are of high quality. These examples should perfectly match the style and format as the desired output, this greatly increases the chance the LLM output will be accurate and relevant [16].

Common pitfalls in prompting include the use of overly broad instructions, mixing tonal or structural styles and the inclusion of personal opinions, which can produce incoherent or hallucinated responses, reducing the quality and trustworthiness of the model's output [73]. LLMs tend to reflect sentiment and affirm assumptions provided by the user, which can compromise objectivity, especially in analytical tasks. Maintaining consistent tone and structural logic is therefore important, particularly when using structured approaches such as CoT prompting.

Lastly, prompt engineering is inherently an iterative process. There is rarely a perfect prompt on the first try; instead, prompting involves repeated testing and refinement based on the model's responses. Observing how changes in structure, wording or context affect the output allows users to fine-tune prompts for both accuracy and clarity [46, 68]. This iterative nature further highlights why prompting is not just a technical trick, but a professional skill that continues to evolve alongside advancements in LLM capabilities.

Following these best practices enhances consistency, reduces hallucinations and maximizes the relevance. It is hard to master all these techniques, which highlights why prompting in combination with prompt engineering is now recognized as a vital skill that continues to evolve.

#### **2.1.4.4 Prompting as a skill**

Once a nascent concept, prompting has rapidly evolved into a critical digital skill in the 21st century, fundamentally reshaping how humans interact with AI systems. With the emergence of advanced LLMs, guiding model behaviour through prompts has become central to AI use, particularly in professional and analytical contexts [49]. This shift has given rise to prompting as a distinct and valuable competence, necessary to translate complex tasks into effective machine-readable instructions. Prompting is no longer a passive interaction but a learned process that demands a combination of technical precision, cognitive structuring and creative articulation. Unlike traditional information literacy, which focuses on evaluating and retrieving static information, prompting requires users to actively construct dynamic input that influences model reasoning and output [46]. As such, it aligns closely with earlier discussed concepts like prompt engineering, which relies on the iterative refinement of inputs to improve relevance, accuracy and transparency. As Federiakin et al. argue, prompt engineering is an emerging skill that supports

both personal adaptability and professional capability in AI-mediated environments [29]. The growing importance of prompting as a professional skill reflects the broader transformation in human-AI collaboration. It is no longer sufficient to understand what an AI can do, users must now know how to instruct it effectively. This competence continues to evolve alongside model capabilities, positioning prompting as a foundational skill for effective AI-supported strategic decision-making.

## 2.2 Strategic planning

Every organization regardless of industry, size or field has a fundamental purpose. This purpose can be to maximize shareholder value, focus on innovation, expand into new markets, contribute to society or enhance environmental sustainability. Large multinational organizations might aim for global market dominance, while start-ups often prioritize agility and growth. However, having a clear purpose is not the full picture; organizations also need to determine how to achieve these objectives effectively [12].

Achieving these objectives requires aligning resources, analyzing market trends, positioning competitively and making informed decisions [12]. The fast pace of technological and market change makes this process increasingly complex and time-intensive [62]. As early as 1986 Orsini underlines this by noting that strategic planning involves multiple factors, including but not limited to intensive data gathering, in-depth analysis, stakeholder meetings, scenario evaluation and report generation [62]. These challenges affect diverse sectors, from healthcare and finance to technology and consulting that remain especially salient today. Their prominence is intensified by the volume and velocity of available data and, increasingly, the rise of AI in the business world.

This study focuses specifically on the management consulting sector, conducting the research within a Big Four consulting firm. In this context, strategic planning stands out as one of the most influential and widely used tools in business, providing a structured approach to navigate uncertainty, optimize performance and maintain a competitive edge in an evolving world [13, 83]. The adoption of strategic planning is especially seen in contemporary organizations [35, 83]. Contemporary organizations are organizations that are decentralized, agile, strive for innovation and engage in teamwork enabling a faster response to market changes due to reduced bureaucracy [74]. The consulting firm under study reflects these characteristics, making it a relevant and representative context for this research.

Within this environment, the growing integration of generative AI, especially LLMs like ChatGPT, presents new opportunities to support strategic planning process and addressing associated challenges. These can generate context-aware insights, synthesize reports and support environmental analyses offers potential efficiency and quality gains in strategic planning activities.

### 2.2.1 Definition and purpose

Strategic planning is a structured process that organizations use to define their long-term goals and determine the best ways to achieve them [13]. This is executed through a process, a process that involves formulating strategies to achieve these goals and effectively allocating resources to ensure success, a task that can be undertaken by management consultants who apply their specialized knowledge and expertise [83, 51]. The appropriate time horizon for strategic planning remains contested in the literature. While some experts contend that the ideal time

range for strategic planning is around 1.5 to 3 years, others argue that the optimal duration is highly dependent on sectoral dynamics and environmental uncertainty [6, 9]. Research shows that during environmental uncertainty managers tend to expand the scope and intensity of strategic planning efforts [9]. For instance, technology-intensive industries often benefit from shorter horizons to remain adaptive to rapid innovation cycles, whereas capital-intensive sectors may favor longer planning periods to accommodate investment recovery and regulatory stability. According to Bryson, strategic planning is essential for organizations to stay aligned with their goals and missions [13]. This is done continuously and iteratively an insight that resonates with earlier discussions on prompt engineering and LLM usage, where iteration and refinement are also critical [13]. Frameworks such as SWOT analysis exemplify this iteration, helping organizations reassess internal strengths and weaknesses alongside external opportunities and threats to evaluate strategies [7]. The SWOT analysis is one of many tools and frameworks within strategic planning. This allows organizations to adapt to changing circumstances and remain relevant in the future [13, 12].

In other words, strategic planning is used when there is need for clarity, direction and focus within an organization. It becomes particularly important when an organization faces significant challenges, such as conflicting demands, resource limitations or changing external conditions [13]. Using strategic planning the organization can identify priorities, align their efforts and develop actionable steps to address these challenges effectively.

A further crucial benefit of strategic planning is its role in improving decision-making [12]. Nutt found that at least half of all strategic decisions fail due to poor decision-making processes [59]. Strategic planning helps mitigate this risk by directing attention to critical issues and challenges, as mentioned above, within the organization, enabling key decision-makers to take well-informed actions.

### **2.2.2 Practitioners of strategic planning**

Mintzberg argues that strategic planning should not be limited to executives and planners, as effective strategy-making benefits from the insights and experiences of individuals across all layers of the organization [55]. Having strategies purely based on executives can result in rigid, overly analytical strategies that are disconnected from existing operational challenges [55]. In contrast, involving middle managers strengthens contextual awareness in defining the right strategy. Positioned between executive leadership and operational teams, they possess both upward access to strategic vision and downward familiarity with practical operational challenges [84]. Moreover, as of now, middle and lower management tend to lack involvement with regards to strategic planning within the organization, while top management is not involved in the overall operations [34]. Gandrita emphasizes that involving everyone in the organization's strategic planning creates a shared understanding of goals and fosters a sense of shared ownership across all levels [34]. Therefore, it essential to not only focus on top management but include all layers of the organization.

In this light, the integration of generative AI tools like ChatGPT holds potential to support more inclusive strategic planning. By lowering the barrier to generating structured input (e.g., through prompting), LLMs can enable non-executives to articulate insights, identify risks or summarize local knowledge in ways that can feed into strategic planning processes.

### **2.2.3 Management consulting**

Strategic planning is important for contemporary organizations, helping them set goals, use resources effectively, adapt to changing markets and increase organizational performance [83, 35]. While strategic planning is commonly associated with corporations seeking competitive advantage, it is also essential within the management consulting sector itself [51]. Consulting firms, like the organizations they advise, must continuously refine their services, develop internal capabilities and adapt to changing business environments to differentiate themselves from established competing companies, such as McKinsey or Bain Consulting Group (BCG), as the industry keeps expanding both geographically and in scope [15]. A study of Kenyan management consultancy firms revealed that strategic planning helps them develop competitive strategies, focus on core competencies and adapt to the environment [51]. Moreover, the Maina study highlights that organizations engage management consultants to address complex management problems, identify and seize new opportunities and guide implementation changes [51]. This two-sided dynamic means that consulting firms not only deliver strategic planning services, they must also actively practice them internally to remain agile and relevant.

Management consulting can be described as either a professional advisory service or a method of providing practical advice for a task where the consultant has no direct control over implementation [47]. According to Whittington consultants are outside strategy advisers in the context of strategic planning, as they provide an unbiased outsider's perspective [82]. Their profession is to provide added value to their clients by leveraging business or IT knowledge to solve management and operational challenges.

However, while consulting firms focus on guiding external organizations, they themselves must also engage in strategic planning. According to Kubr, management consulting is "an independent professional advisory service assisting managers and organizations to achieve organizational purposes and objectives by solving management and business problems, identifying and seizing new opportunities, enhancing learning and implementing changes" [47]. This definition underscores that consultants help organizations develop actionable strategies, a core function that mirrors the essence of strategic planning itself. As consultants guide other businesses in defining goals and structuring their paths to success, they must also apply similar principles to ensure the growth and sustainability of their own firm, especially in a fast-changing, tech-driven landscape. This is realized by implementing strategic planning in the organization [51, 71].

As discussed in Subsection 2.2.2, the strategic planning process benefits from input across all layers of the organization. Within consulting firms, this includes everyone from junior analysts

to senior partners. Here the integration of generative AI hold particular relevance. By supporting ideation, generating tailored analyses and enabling fast access to contextual insights, LLMs can enhance consultants' ability to contribute at every level [44]. Moreover, as consultants are often early adopters of digital tools, their use of LLM for internal strategy and client analysis illustrates how prompting competence can evolve into a core consulting skill.

#### **2.2.4 Process of strategic planning**

As mentioned above, strategic planning is a structured process designed to define long-term goals, formulate strategies and allocate resources efficiently. Within the management consulting sector, this process is not only essential for guiding client organizations, but also for ensuring that consulting firms themselves remain competitive, agile and well-positioned in an evolving industry landscape [47].

There are various definitions of the strategic planning process. Mintzberg characterizes it as a highly structured, formalized and bureaucratic process [54]. According to him, the planning model involves setting objectives, conducting internal and external analysis, evaluating the chosen strategy, operationalizing it and afterwards keeping control of it. However, he argues that such a bureaucratic and rigid approach might hinder flexibility and responsiveness within the organization [55]. Schendel and Hofer proposed a nearly identical process, but differentiated it by explicitly including a strategy formulation phase, resulting in a total of six distinct steps [71]. Meanwhile, Ketokivi and Castañer describe the process of strategic planning as activities conducted periodically, such as budgeting, resource allocations and annual assessments, framing the process as both cyclical and embedded in organizational operations [45]. Bryson describes strategic planning as a structured yet iterative process involving goal-setting, strategy formulation, resource allocation and regular assessments, ensuring organizations stay aligned with their mission while adapting to internal and external changes [12].

This study will adopt the strategic planning model proposed by Schendel and Hofer as our primary framework, as it provides a clear and structured approach. The strategic planning process used in this research is presented in Figure 2.9. This figure shows that the strategic planning model is a linear yet iterative process. The cyclical nature, represented by the node "iterative feedback", shows that the strategy control & monitor step provides iterative feedback to all preceding steps. To further enhance this model's applicability, it will be supplemented with insights from contemporary academic literature, allowing other researchers to build upon and refine it.

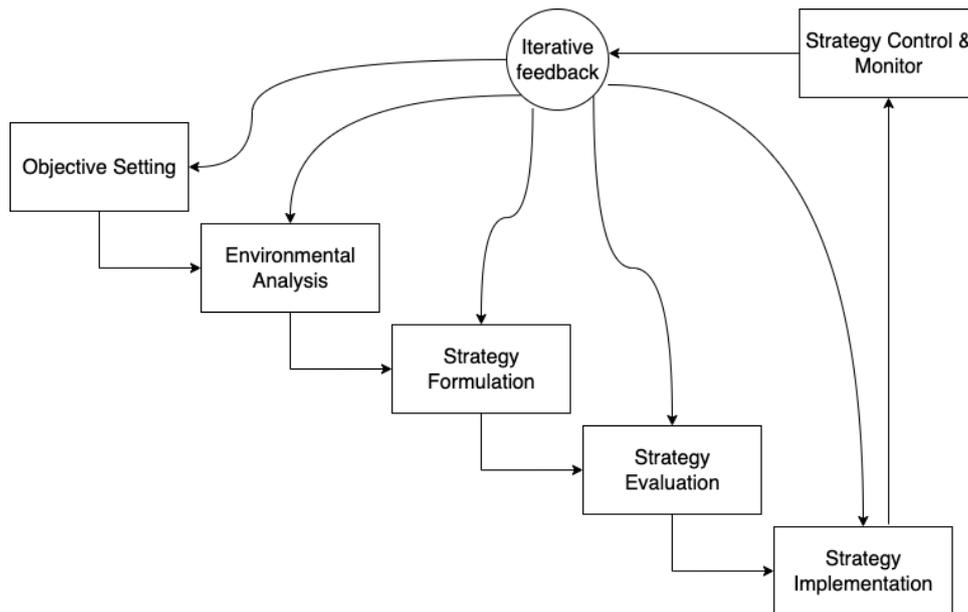


Figure 2.9: The six-step strategic planning process

The first step is determining organizational objectives. These goals can be measurable, such as revenue growth or increased market share, but must be framed carefully. However, this must be approached with caution because an over-reliance on numerical targets can blur the line between setting objectives and actually developing a strategy [54]. While goals are integral, they should not overshadow the strategic process; the emphasis must remain on how to achieve these objectives, rather than solely on meeting the numeric benchmarks. Objectives should guide action, not become the sole metric of success.

The second step, which is quite extensive, is conducting an environmental analysis to evaluate both external and internal influences on the organization that could affect a company's success [13, 83, 71]. Various analytical tools exist for this step. One common framework is the SWOT analysis, which identifies strengths, weaknesses, opportunities and threats of an organization [7]. Another widely used framework is PESTLE analysis (political, economic, social, technological, legal and environmental), which can be employed alongside SWOT to ensure a more comprehensive environmental evaluation [83, 2]. Stakeholder and competitor analyses, such as Power-Interest Matrix or Porter's Five Forces, are also critical for understanding the broader environment, as they have a major effect on the organization [64, 54, 13, 52]. Combining all these tools produces more accurate and powerful results and strategic decisions [7]. The effectiveness of these tools depends on identifying the right factors to analyze, such as political, economic and technological trends, competitive pressures or stakeholder interests. Overlooking a critical factor can be more detrimental to an organization than inaccurately forecasting its impact. Therefore, selecting the most appropriate tool or a combination of tools, based on the specific strategic context is essential [71]. These tools, especially when combined, play a crucial role in strategic planning by offering a structured approach to analyzing internal and external influences. Given the increasing integration of AI in business processes, it is

important to examine how generative AI can enhance these tools [63]. Research has already shown that generative AI can improve efficiency and decision-making by automating data analysis, identifying patterns and generating insights that might be overlooked by traditional methods [69, 44]. Particularly in management consulting, where time-constrained analyses are common, LLMs like ChatGPT can augment this phase by enabling consultants to prompt effectively to enhance strategic planning work.

The third step is strategy formulation, where insights gained from internal and external analyses guide the development of strategies. This phase positions the organization strategically within its environment to achieve established objectives effectively [71]. The strategy formulation involves clear steps, according to Schendel and Hofer: (1) problem identification, (2) problem analysis, (3) generation of alternatives, (4) evaluation of alternatives and (5) selection of the best option [71]. This structured approach is aimed at solving organizational challenges systematically. However, they acknowledge that strategy formulation can also be informal or implicit, particularly in smaller or entrepreneurial organizations, where strategy may emerge organically from the founder's intuitive understanding of the organization and its environment rather than through formal analysis. Since management consultants typically work with larger organizations where formal strategy formulation is standard, we include the environmental analysis as a key component of the strategic planning process. In this phase, LLMs can also play a supportive role, enabling consultants to quickly prototype multiple strategies and compare them through iterative prompting.

The fourth step involves strategy evaluation. Here, the organization assesses whether the proposed strategy under the given environmental conditions leads to the desired objectives [71]. According to Schendel and Hofer, evaluation requires answering two fundamental questions: first, whether the existing strategy adequately meets organizational goals and second, whether any proposed alternative would perform better given anticipated environmental circumstances [71]. Techniques to evaluate this are risk analysis, early performance metrics, return on investment measures and tools for testing strategic consistency and feasibility [71]. Another proposed method to evaluate the strategy is through value creation; will the value of the company increase by this strategy [54]? Lastly, other key evaluation criteria include consistency (internal coherence), consonance (external fit), advantage (competitive edge) and feasibility (realistic within resource limits) [54]. LLMs can be prompted to assess strategy feasibility or alignment based on predefined evaluation criteria, offering rapid assessments or generating structured documentation to support decision-making. In addition, various strategic scenarios can be quickly created and compared or evaluated by LLMs. Moreover, preliminary risk analyses can be conducted, enabling consultants to identify potential weaknesses or trade-offs before full implementation. This not only accelerates the evaluation process but also enhances analytical rigor by introducing alternative perspectives and data-informed reasoning into strategic discussions.

The fifth step is strategy implementation, where the chosen strategy is converted into actionable tasks and executed throughout the organization [54]. According to Schendel and Hofer, implementation primarily involves administrative tasks, including aligning organizational

structure, processes and behaviour's with strategic objectives [71]. Key factors are to determine clear roles and responsibilities and accurately allocating resources. Consultants often play a facilitative role in ensuring that the transition from planning to action is executed without loss of strategic intent [47].

Finally, the last step involves control and monitoring. To control and monitor the strategy, key performance indicators (KPIs) are determined and regularly checked to see whether these indicators are being met [54, 71]. These KPI's can be derived from the early performance indicators. As the process is not linear, rather continuous and iterative, insights gained from this evaluation may also feed back into earlier steps [12]. For example, reassessing the mission or refining objectives to ensure continued alignment with the organization's evolving environment.

To sum up, strategic planning proceeds like a structured journey: the organization first clarifies its main concrete, measurable goals (objectives). Next, it analyzes both internal and external factors (environmental analysis) to gain a clear situational overview. Building on these insights, it design appropriate strategies (strategy formulation), allocates resources and evaluates the feasibility of each plan (strategy evaluation). Finally, the chosen strategy is implemented (strategy implementation) and regularly monitored (implementation and control) to ensure ongoing alignment with the organization's objectives.

## **2.2.5 Strategic planning activities**

Having established the importance, purpose and general process of strategic planning, as well as briefly introducing its activities, we now turn to a more detailed examination of those activities. As discussed above, the strategic planning process involves six steps, each containing operational tasks that guide an organization from initial goal-setting to execution and continuous improvement. These encompass extensive analytical tools such as PESTLE and SWOT analyses, Porter's Five Forces, stakeholder and risk assessments and resource analysis. Also critical are performance monitoring, the determination of Key Performance Indicators (KPIs) and operational tasks like generating reports for management and facilitating brainstorming sessions [83, 71, 12]. These activities together form the practical foundation of strategic planning. The following subsections dive deeper into some of the key activities practiced by management consultants within the strategic planning process.

### **2.2.5.1 SWOT analysis**

SWOT analysis, established in the 1960s and popularized in the 1970s, evaluates an organization's strengths, weaknesses, opportunities and threats to inform strategic decision-making [7]. It has been widely adopted as a strategic planning tool both by academics and practitioners [7]. Key activities include identifying internal strengths and weaknesses through financial performance reviews, operational assessments and surveys [7]. External analysis examines market trends, competitor strategies, regulatory changes and technological advancements.

In consulting settings, LLMs can enhance SWOT analysis by assisting with automated data summarization, extracting patterns from documents or meeting notes and synthesizing relevant external developments. When properly prompted, such tools offer consultants a head start in drafting SWOT matrices or identifying emerging threats and opportunities.

#### **2.2.5.2 PESTLE analysis**

Evolving from earlier PEST frameworks, the PESTLE analysis evaluates the external macro-environmental factors that affect an organization, namely political, economic, social, technological, legal and environmental factors. Its origins can be traced back to the 1960s, with foundational work such as Aguilar's 1967 study on scanning the business environment [2].

In today's volatile and fast-paced world, the technological dimension of PESTLE has become particularly salient. Organizations increasingly face technological uncertainty, which is the difficulty of predicting not just when innovation will occur, but how emerging technologies will impact operational processes, competitive dynamics and strategic positioning [25]. With disruptive advancements such as generative AI, the question is no longer if change will occur, but how and to what extent such technologies will reshape entire industries. In this regard, generative AI assumes a dual role within the PESTLE framework: both as a subject of analysis (technological disruption) and a tool that can support the analysis itself.

#### **2.2.5.3 Stakeholder analysis**

According to Freeman, a stakeholder is "... any group or individual who can affect or is affected by the achievement of the organization's objectives" [32]. Brainstorming with leadership roles, reviewing previous projects and analyzing documents, such as organizational chart can all identify potential stakeholders. It is important to consider whether to include internal stakeholders (e.g., employees, management, partners) or external stakeholders (e.g., customers, regulators, media), as this decision guides the analysis. A well-known model to map stakeholders is the Power-Interest Matrix, developed by Mendelow. This model is a 2x2 matrix that maps stakeholders based on their level of power and interest, helping organizations prioritize engagement strategies [52]. Examining stakeholders clarifies their expectations and helps define the organization's purpose collectively. Moreover, in many cases organizations require stakeholder approval for certain decisions. Therefore, it is essential to understand stakeholders and identify key influencers, decision-makers and affected parties both within and outside the organization.

LLMs can support this process by generating draft stakeholder profiles using role-based prompting or by analyzing sentiment from meeting transcripts and notes. These outputs offer consultants a valuable starting point for deeper stakeholder engagement.

#### **2.2.5.4 Porter's Five Forces**

Developed by Michael Porter in 1979, this model assesses the competitive intensity of a market by examining five key forces that influence industry profitability [64]. The five forces are:

threat of new entrants, bargaining power of suppliers, bargaining power of customers, threat of substitute product or services and the industry; jockeying for position among current competitors [64]. These forces have the following effect on organizations: suppliers and buyers can influence pricing and contract terms, new entrants and substitute offerings can unsettle the market and ongoing battles among established rivals intensify industry competition [48, 64]. Therefore, it is essential to understand them. Porter's five forces focus on the industry an organization is in, as opposed to PESTLE that focuses on macro-environment influences [48]. In order to successfully execute this tool, several analytical tasks must be performed. Industry data collection is essential for identifying all five forces and can be achieved through industry reports, financial statements and market research. To enhance the analysis, SWOT analysis can be integrated to align internal strengths and weaknesses with Porter's model [7]. Additionally, a stakeholder analysis can support this process by constructing a stakeholder matrix, providing a clearer visualization of interactions between suppliers, buyers and competitors. This, in turn, helps in identifying potential substitutes and assessing their impact on industry dynamics. A comprehensive application of Porter's Five Forces Model enables firms to not only assess industry dynamics but also proactively shape their competitive positioning by mitigating risks and leveraging opportunities for sustained profitability [64].

#### **2.2.5.5 Operational tasks**

Strategic planning is supported by operational activities that ensure decisions translate into actionable outcomes. One such activity is performance monitoring, typically guided by KPI's. KPI's help with this, as they provide an indicator of how the company is performing at that specific time. KPI's can be set at various levels within the organization, such as market share growth for the entire company or utilization rate as a team-specific KPI. The latter is particularly important in the management consulting sector, as the revenue model of these organizations relies on selling expertise and consequently, consultant time or people. A high utilization rate indicates that consultants are effectively engaged in billable client work, maximizing revenue and operational efficiency. Another core task in strategic planning is resource allocation. In order to make sure the mission statement is achieved, resources should be allocated according to what is required. This can be done through a resource planner that determines which department or team receive what resources. There are also various operational tasks that support the higher levels in an organization. Junior consultants for example, generate reports that in turn are sent to their managers, which use them to analyze internal or external data to support with strategic decision-making.

The activities discussed above form the practical core of strategic planning. By breaking down abstract goals into concrete, repeatable tasks organizations can plan, act and adapt more effectively in fast-changing markets. This activity-level perspective is particularly important for assessing how generative AI, especially through tools such as Large Language Models (LLMs), can play a role. Whether it's speeding up environmental analysis, drafting reports or mapping stakeholder concerns, these tools offer new ways to support and possibly reshape

how strategy gets done. A deeper understanding of these individual activities enables a more nuanced evaluation of where and how AI-driven tools can contribute meaningful improvements to strategic decision-making.

# Chapter 3 Methodology

This chapter outlines the research methodology used in this research. For reference, the research question is presented down below:

**RQ.** *What are the benefits and challenges of using Generative AI to support strategic planning activities in management consulting?*

To answer the research question effectively, an exploratory sequential mixed methods approach has been adopted. This approach combines both qualitative and quantitative methods and is finalized by a practical demonstration. The qualitative phase consists of exploratory expert interviews, aimed at building upon the existing literature to gain a deeper understanding of the specific activities involved in strategic planning. The findings from this phase combined with the literature review inform the construction of a quantitative questionnaire. Finally, a practical demonstration is presented to illustrate generative AI in practice on strategic planning activities. This chapter proceeds as follows: first, the overall research design and phases are introduced. Then, data collection procedures, sampling techniques, data analysis methods and validation & reliability are discussed separately for both the qualitative and quantitative components. The chapter concludes with an explanation of the practical demonstration.

## 3.1 Research design

To investigate the benefits and challenges of using generative AI in strategic planning activities within management consulting, this study employs a sequential mixed methods design. A mixed methods design is when two different research methods are used within the same research [38]. Combining research methods enables a more comprehensive analysis than using either qualitative or quantitative approaches alone, thereby offering richer and more nuanced insight to the research problem [40]. The integration of both approaches leverages the respective strengths of each, providing depth and breadth to the findings [30].

There are various design levels in mixed methods research, ranging from basic to advanced frameworks [30]. This study adopts an exploratory sequential design, in which qualitative data collection precedes and informs the quantitative phase. This choice is motivated by the need to first gain in-depth insights through expert interviews, which then guide the development of a structured survey instrument. According to Bryman, one rationale for mixed methods is instrument development, where qualitative research is used to inform the design of survey questions and variables [11]. This aligns closely with the aim of this research: to uncover themes not yet fully formalized in the literature and to test these at scale. Therefore, this mixed methods approach is a perfect fit for achieving the research goal.

In addition, this rationale follows Bryman's framework, which highlights "completeness" and "development" as valid justifications for mixed methods [11]. In this study, interviews serve to develop the questionnaire and ensure completeness by capturing perspectives that may not yet be fully formalized in the literature. Thus, the research design follows a qual → QUAN structure, where the qualitative phase has a formative role, but priority is given to the quantitative phase, which ultimately answers the main research question.

Nonetheless, mixed methods research also presents several challenges. One major challenge, particularly for junior researchers, is selecting an appropriate design, as this choice significantly influences the research outcomes [21]. To address this, the exploratory sequential design was deliberately chosen, as it aligns well with the research objective. The activities of strategic planning will be explored first, then examined where generative AI can support these activities. Because the design requires gathering and analyzing two distinct datasets, it typically demands more effort and coordination than a single-method study [21]. This will be addressed by limiting the number of interviews while still ensuring sufficient depth and breadth to inform the subsequent survey design.

In addition to the qualitative and quantitative phases, this study incorporates a practical demonstration to illustrate how generative AI can be applied in strategic planning activities. According to Creswell, including visual and applied elements in mixed methods research can enhance interpretability, particularly when exploring emergent technologies or new practices [17]. While the exploratory sequential mixed methods approach forms the methodological backbone of this thesis, a practical component enriches the research by providing a concrete and experiential understanding of how LLMs can operationalize theoretical insights. The format is inspired by the work of Jarco, who similarly employed a mixed methods design and concluded with a visual presentation of LLM outputs to enhance interpretability and insight [41]. To ensure transparency and replicability, full prompt designs and corresponding model outputs will be included in the appendix.

## **3.2 Qualitative phase**

The research begins with the qualitative phase. This will be conducted through exploratory expert interviews. The expert interviews are primarily aimed at uncovering expert perspectives on the activities of strategic planning and how generative AI can have an impact on these activities. As Creswell emphasizes, qualitative inquiry is well-suited for studies seeking to explore complex, under-researched phenomena, such as the specific activities of strategic planning, through engagement with participants [17]. As the existing literature lacks robustly describing the activities of strategic planning, it is crucial to create a more complete understanding of what activities are used in strategic planning and how this can be influenced by generative AI. Semi-structured interviews are selected as the primary method for qualitative data collection, as they allow for flexibility while still providing comparability across participants.

However, qualitative research is not without its pitfalls. Common criticisms include potential researcher bias, insufficient scientific rigor and a lack of transparency in the analytical process [58]. To mitigate these concerns, several measures are implemented. First, all interview questions will be open-ended rather than closed-ended, allowing for the emergence of novel insights and helping to reduce researcher bias [42]. Second, the expert interviews will follow a semi-structured format with a pre-developed interview guide. This approach ensures consistency across all interviews, thereby enhancing the reliability of the results and contributing positively to the study's rigorousness [8].

### **3.2.1 Data collection**

Data will be collected in the qualitative phase through expert interviews, which are audio-recorded. These are exploratory in nature, as the goal is to gain a deeper understanding of the subject. As mentioned above, the interviews will follow a semi-structured design. Each interview follows a flexible interview guide that ensures key topics are covered but leaves spaces to pursue emerging insights. This format is well-suited for investigating complex and emergent topics, as it offers flexibility and enables in-depth probing where appropriate [33]. Given the complexity of strategic planning and the limited prior research on its activities, particularly in relation to the emerging topic of generative AI, the flexibility to ask follow-up questions makes this method well-suited for this research. The interview guide will be followed as closely as possible to ensure the reliability of the results, while considering timing and allowing for natural variation in the depth of responses [8]. The interview design aims to explore key themes and deepen the understanding of strategic planning activities. In addition, it investigates how experts perceive the impact of generative AI on these activities. To ensure clarity and relevance, the questions were pre-tested with an expert in the field.

### **3.2.2 Sampling technique**

The sampling strategy for the qualitative phase targets professionals in the management consulting sector, specifically those employed by large international consultancy firms and a major enterprise software provider. These organizations operate at the intersection of technology and strategy, making them highly relevant to the research topic. The selected consultants vary in seniority and specialization to ensure diverse perspectives. Rather than aiming for statistical representation, the goal was to obtain deep, experience-based insights from professionals embedded in the field of interest. This purposive sampling approach supports the exploratory nature of the study and provides a solid empirical foundation for informing the subsequent quantitative phase.

As noted in Section 3.1, mixed methods research is time-intensive. Therefore, the number of interviewees is deliberately limited. The goal is to gain a deeper understanding of the research subject, when this is reached there is theoretical saturation. When the topic is fully explored and more data does not result in additional theoretical insights, theoretical saturation is achieved [39]. Research suggests that between one and six interviews are often sufficient for identifying

major themes in homogeneous expert samples [36]. Accordingly, this study will conduct six expert interviews.

### **3.2.3 Data analysis**

In order to analyze the data from the qualitative phase, the expert interviews will be transcribed. Transcriptions are essential for performing a thorough and academically sound analysis and are a central practice in qualitative research [20]. The transcriptions will serve as the primary data for the subsequent content analysis to identify novel strategic planning activities. As this research is exploratory and primarily focused on instrument development, the initial phase of content analysis will be approached with a broad lens to identify strategic planning activities. All strategic planning activities identified by the interviewees will be systematically documented in a list.

Additionally, expert perceptions of where generative AI can add value in strategic planning will be analyzed using a Modified Borda Count (MBC). Each respondent is asked to rank their top three strategic planning steps (out of six) where they expect generative AI to provide the most benefit. In line with MBC logic, ranks were converted to points: 3 for first place, 2 for second and 1 for third, while unranked steps received 0 points. This method accounts for partial rankings and allows for aggregating ordinal data into a group-level prioritization. The MBC approach is well-established for processing incomplete preference lists while preserving the proportional influence of each vote [27]. This will then inform the focus of the quantitative phase, indicating which steps of the strategic planning process will receive more detailed attention in the instrument.

### **3.2.4 Validation & reliability**

Ensuring trustworthiness is paramount in qualitative research. Reliability and validity, though interpreted differently in qualitative inquiry, are addressed through several procedural measures.

Reliability is enhanced by the use of a semi-structured format and pre-developed interview guide, which ensures that a comparable set of topics is covered across all expert interviews [8]. Furthermore, all interviews will be audio-recorded and transcribed [20]. This creates an auditable data trail to enhance transparency in the analysis phase.

Validity is maximized through two key strategies. First, purposive sampling ensures that participants are deeply embedded experts, providing high-quality experience-based insights directly relevant to the research topic. Second, the initial content analysis will use an exploratory, broad lens ensuring that novel and emerging strategic planning activities are accurately captured as they are voiced by the experts, thereby maximizing the authenticity of the findings [17].

### 3.3 Quantitative phase

To conclude the mixed methods approach, a structured questionnaire will be developed for the quantitative phase. This survey translates the insights gathered during expert interviews into measurable constructs, thereby aligning with the principles of an exploratory sequential mixed methods design [40, 17]. Whereas the qualitative work identified activities per strategic planning step that are performed by consultants, the quantitative phase measures how large the unrealized potential is and for whom that potential is most noticeable. In other words, the primary objective of this phase is to identify the most and least promising activities for using generative AI in strategic planning activities among a broader population of management consultants. The resulting gap profile will determine the activity for the practical demonstration phase discussed in Appendix C.

To operationalize this objective, the survey adopts the gap-analysis logic, as derived from Dybå & Moe and further inspired by a more recent paper by Stettina & Heijstek [26, 75]. Dybå & Moe discussed that for each item in their paper, respondents were asked two questions: the current satisfaction and future importance, which were subsequently compared to identify a gap. This research follows the same methodology with a minor adjustment. Consistent with the logic of the aforementioned gap analysis, every strategic planning activity is measured twice, using a paired-item format. Respondents are asked two questions per activity:

1. **Current satisfaction** - 'I am satisfied with the current usefulness of Generative AI in supporting *ACTIVITY NAME*.'
2. **Future importance** - 'I believe Generative AI will be important in the future for *ACTIVITY NAME*.'

Both judgments use an identical seven-point Likert scale: *Not at all* = 1, *To a very small extent* = 2, *To a small extent* = 3, *To a moderate extent* = 4, *To a large extent* = 5, *To a very large extent* = 6 and *To an extremely large extent* = 7. To ensure higher sensitivity and more accurate capture of respondents' subjective evaluations, crucial for identifying subtle differences in perception, a 7-point Likert scale was used. Prior research has shown that 5-point scales are more likely to cause response imprecision, i.e., participants attempting to answer between two discrete options, thereby reducing measurement precision in electronically-distributed surveys [31]. The numerical nature of the Likert scale allows for the direct calculation of the gap score, computed as:

$$\text{Gap}_i = \text{Future Importance}_i - \text{Current Satisfaction}_i \quad (3.1)$$

This equation quantifies the unrealized potential of generative AI for activity *i*. Gap values range from -6 to +6 Likert points, with large positive values indicating high opportunity and negative values suggesting limited added value.

The questionnaire retains the six-step planning model discussed in Section 2.2.4, but detailed measurement is limited to two steps, in line with the interview-based Modified Borda Count. Their respective activities are then presented in the survey with the paired satisfaction-importance items as described above. To anchor the gap values in broader attitude theory and enable better insights into a respondents' character, four single-item constructs are inserted at the beginning of the questionnaire:

**Perceived usefulness** "I am satisfied with the current usefulness of Generative AI in my professional work."

**Perceived AI importance** "I think Generative AI will be important in the future in my professional work."

**Perceived ease of use** "I would find Generative AI easy to use in my professional work."

**Perceived risk** "I feel that using Generative AI at work could pose potential risks."

Demographic and control variables (e.g., age, seniority, specialization, frequency of generative AI use and strategic planning work) follow, allowing for subgroup analysis. All items employ the same seven-point format, thereby avoiding scale-switching errors and permitting direct arithmetic operations on the gap variable. The survey is structured in the following order: (1) initial attitude constructs, (2) demographic and control variables, (3) paired satisfaction-importance items for selected strategic planning activities and (4) section on prompting.

The remainder of this chapter sets out how the survey will be distributed and collected, who will be sampled, how the responses will be analyzed and how the instrument's reliability, validity and potential biases are addressed.

### 3.3.1 Data collection

Data for the quantitative phase will be collected through a self-administered online survey, completed by management consultants. Qualtrics will serve as the platform to design and distribute the questionnaire. An online format was chosen, because it allows rapid distribution to many consultants with minimal administrative efforts [79, 5]. Given the digital literacy of the target group, an online format is especially appropriate [5].

To promote data accuracy and response validity, confidentiality and anonymity will be guaranteed throughout the survey process [3, 56]. Ensuring respondent privacy helps to reduce self-reporting and social desirability biases, which are common concerns in survey-based research.

However, online surveys are not without limitations. One key concern is survey fatigue, which can occur when questions require long-term memory recall, detailed calculations or when survey patterns are not straightforward [76, 79]. Given the benefit of a larger sample size for this research, survey fatigue will be carefully mitigated. To mitigate this, a pilot test will be conducted with an expert in management consulting, who will evaluate clarity, cognitive load

and signs of fatigue. Additionally, consistent use of the seven-point scale throughout reduces switching costs and keeps respondents engaged. These measures aim to maintain respondent engagement and reduce dropout rates. After final adjustments, the survey will be distributed to the target population.

### **3.3.2 Sampling technique**

The target population consists of management consultants within the advisory department of a Big Four consulting firm, encompassing more than 2000 individuals. The goal is to obtain at least 100 responses, which would require a 5% response rate. As argued in Section 2.2.2, strategic planning is a multi-layered organizational process that requires the involvement of individuals at all hierarchical levels. Consequently, the decision to include consultants from all seniority levels, ranging from consultants to partners and across various functional disciplines ensures a diverse and representative set of perspectives within the organization.

Random sampling in online surveys is challenging due to the lack of a clear framework for traditional probability sample selection [5]. To address this, Ball recommends that researchers define a target population that is bounded, well-specified and relevant to the research objective [5]. Following this guidance, purposive sampling technique is adopted to identify the target group; management consultants within the advisory department of a Big Four consulting firm. Given their direct engagement with strategic planning and potential exposure to generative AI tools, this group is particularly well-positioned to inform the research question.

### **3.3.3 Data analysis**

To ensure the rigor and validity of the quantitative data analysis, several preliminary and analytical steps will be undertaken. Initial data processing will involve screening responses for quality and completeness. Responses failing predefined attention check questions will be systematically identified and excluded, a crucial step to mitigate the potential impact of inattentive responding on the validity of findings [19]. The dataset will also be examined for missing values.

Following data cleaning, descriptive statistics will be calculated and presented to provide an overview of the sample characteristics and control variables. Categorical demographic variables, such as age, seniority, specialization and job title will be summarized using frequency distributions and visualized using bar charts to provide a better understanding of the data [43]. Additionally, the distribution of attitude items (perceived usefulness, perceived ease of use, perceived risk, perceived AI importance) and AI usage will be analyzed by computing means. Histograms will help visualize their distribution, allowing for an assessment of skewness and kurtosis.

### 3.3.3.1 Gap analysis

To assess the discrepancy between perceived future importance and current satisfaction with generative AI for specific strategic planning activities, a gap analysis will be conducted. This analysis will be presented in a table displaying, for each activity, the mean and standard deviation of current satisfaction and future importance, the resulting gap score and its standard deviation. Although the primary presentation will focus on average gaps across all respondents, the analysis is fundamentally based on the paired nature of the importance and satisfaction ratings.

- **Gap calculation for activities:** For each activity, the gap score for each respondent will be calculated by subtracting the satisfaction rating from the importance rating. These individual gap scores will then be averaged across all respondents to obtain the mean gap score for the activity. This is calculated as follows:  $\bar{g}_i = \frac{1}{n_i} \sum_{r=1}^{n_i} (I_i^r - S_i^r)$ , where  $I_i^r$  and  $S_i^r$  are the importance and satisfaction ratings of respondent  $r$  for activity  $i$ , and  $n_i$  is the number of paired responses. When all pairs are complete,  $\bar{g}_i$  equals  $\bar{I}_i - \bar{S}_i$  (with  $\bar{I}_i$  and  $\bar{S}_i$  computed on the same paired set). This will provide the average difference between importance and satisfaction for each activity across all respondents.
- **Aggregation within strategic planning steps:** The activity-level gap scores (calculated for each respondent and then averaged) will be averaged within each strategic planning step (Environmental Analysis and Strategy Evaluation) to obtain a mean gap score for each step ( $\bar{g}_{EA}$ ,  $\bar{g}_{SE}$ ). An overall mean gap score ( $\bar{g}_{AL}$ ) will also be calculated by averaging the activity-level gap scores across all activities.

### 3.3.3.2 Software

All statistical analyses will be performed using a combination of the built-in data analysis tools available within Qualtrics and the statistical programming language R. Qualtrics will primarily be used for initial data screening, cleaning and descriptive statistics. R will be employed for making the figures.

### 3.3.4 Validation & reliability

Reliability and validity of the survey and its findings were strengthened through various techniques. These include ensuring respondent anonymity and confidentiality, using clear and well-structured survey items and purposive sampling of a highly relevant population [5, 3]. Furthermore, the survey instrument was grounded in both primarily the findings from the qualitative phase and the literature, enhancing construct validity by ensuring that the items are aligned with the theoretical constructs relevant to the research objective [17]. These measures mitigate common forms of bias such as self-reporting bias, non-representative sampling bias and social desirability bias.

Several additional measures were taken to further strengthen the validity of the survey. Clearly stating the research objective at the beginning of the questionnaire helped respondents

understand the scope and purpose of the study, which contributes to improved response quality [5]. Although reverse-worded items are often used to detect response biases and enhance construct validity, they were deliberately excluded in this study [22]. Given the professional nature of the target population and the cognitive complexity of the constructs, reverse-worded items may have introduced confusion or increased measurement error, as has been observed in high-cognition survey contexts [22]. Instead, the survey emphasized clarity in item wording and logical flow to maintain reliability and participant engagement. Additionally, two instructional attention checks items were embedded in the survey to identify inattentive respondents and further enhance data quality [61].

Nevertheless, it must be acknowledged that this phase remains subject to a degree of subjectivity, due to the self-reported nature of the data. While the design includes several safeguards to enhance validity and reduce bias, these limitations should be considered when interpreting the results.

### **3.4 Practical demonstration**

To complement the exploratory sequential mixed methods design and enhance the interpretability of the findings, this study includes a practical demonstration of generative AI in action. This component aims to bridge the gap between theoretical insights and applied practice by showcasing how a generative AI model, such as OpenAI's ChatGPT, can be used to support key activities within the strategic planning process in a management consulting context. The demonstration contributes to the overall research aim by providing concrete examples of how generative AI tools can be operationalized in a professional setting, based on scenarios derived from both the literature and empirical phases of this study.

ChatGPT was chosen, because the researcher has extensive experience interacting with it across various contexts, has access to the premium version of the tool and is widely adopted within the industry [44]. To optimize the outputs, a set of prompting strategies explained in Subsection 2.1.4.2 will be employed to optimize the outcome [70]. The latest version of ChatGPT will be used: ChatGPT-5.

To ensure transparency and replicability, all prompts and their corresponding outputs will be included in the appendix. This approach not only enhances methodological rigor but also enables future researchers to replicate or build upon this work. The demonstration is presented in Appendix C, to keep the main focus of the thesis on answering the research question through the qualitative and quantitative phase.

# Chapter 4 Results

This chapter presents the empirical findings of the research, structured in accordance with the two-phase exploratory design. The goal is to provide insight into how the impact of generative AI is currently perceived within strategic planning activities in the context of management consulting. The chapter is divided into two main parts: the qualitative results of exploratory expert interviews and the data analysis of the quantitative survey. First, the qualitative results are presented in a framework-guided content analysis. Second, the questionnaire are presented results are presented descriptively.

## 4.1 Qualitative results

### 4.1.1 Participants' demographics

A total of six semi-structured expert interviews were conducted. In line with the sampling criteria outlined in Subsection 3.2.2, all participants are management consulting professionals working in large, internationally active firms that integrate strategy and technology advisory services. The group spans multiple seniority levels, ranging from consultant to partner and covers a mix of functional specializations. This purposive composition was selected to ensure high-quality, experience-based insights, aiming to gain a deeper understanding of the activities involved in strategic planning and how participants perceive the value of generative AI within these specific activities.

# Interviewee	Function	Years of Experience	Company	Specialization	Duration (min)
Interviewee 1	Manager	19	KPMG	Co-lead AI	56
Interviewee 2	Partner	30	KPMG	Sales	52
Interviewee 3	Director	25	KPMG	Supply chain & procurement	65
Interviewee 4	Associate Partner	12	SAP	Discrete industries & energy	61
Interviewee 5	Consultant	1	KPMG	Supply chain management	49
Interviewee 6	Partner	35	KPMG	IT enabled business transformations	20

Table 4.1: Descriptive characteristics and interview duration of the interviewees

The qualitative phase serves two main objectives, which also form the basis for the design of the subsequent survey instrument. The first objective is to identify all the activities management consultants perform for strategic planning. Strategic planning is then subdivided into the six-step model: 1) Objective Setting (OS), 2) Environmental Analysis (EA), 3) Strategy Formulation (SF), 4) Strategy Evaluation (SE), 5) Strategy Implementation (SI) and 6) Strategy Control & Monitoring (SCM). For each step, activities are identified and determined based on the results

of the interviews. An overview will be presented, in the form of a list, illustrating the activities that are linked to the specific activities.

The second objective is to construct a matrix to assess which strategic planning steps are perceived as most suitable for the application of generative AI. This matrix is based on responses to the following interview question: “Can you provide a top 3 of the steps where generative AI can have a positive impact on the respective activities and explain for each step why you chose it?” This objective serves to prioritize the planning steps, allowing for a targeted selection of the most prominent steps to be evaluated more specifically in the survey phase.

### 4.1.2 Content analysis

All interviews were transcribed and systematically reviewed to extract concrete activities mentioned by the expert. These activities identified by the expert interviews are then mapped onto the six steps of the strategic planning model presented in Figure 2.9. Table 4.2 lists the concrete activities per step. The right-hand column shows the interviewee ID(s); see Table 4.1 for the ID-to-interviewee information. Items are grouped by step for clarity, with close synonyms merged. The interviewee IDs just show who mentioned each item; they don’t show agreement or how important it is. In Section 4.1.3, the steps most suitable according to the experts are selected and carried forward into the quantitative phase to test and generalize these exploratory findings.

Table 4.2: Strategic planning activities identified from expert interviews, grouped by planning step.

Strategic planning step	Activity	Interviewee
Goal Setting	Define goals based on data analysis.	1, 2, 3
	Strategic brainstorming and idea generation.	2, 3, 4, 5, 6
	Decision-making on objectives.	3
	Align objectives with organizational vision.	1
Environmental Analysis	Conduct SWOT analysis.	2, 3, 4, 5
	Conduct PESTLE analysis.	4
	Conduct Porter’s Five Forces analysis.	4
	Perform market analysis.	1, 2, 3, 4, 5, 6
	Perform competitor analysis.	3, 4, 5, 6
	Perform stakeholder analysis.	1
	Conduct organizational maturity / AS-IS analysis.	4, 6
	Brainstorm about client/competitor positioning (“market edge”).	2
	Continuously scan technological developments.	4
Strategy Formulation	Define strategic positioning.	3, 4
	Prioritize strategic alternatives.	2

Continued on next page...

Strategy Formulation

**Table 4.2 – continued from previous page**

<b>Strategic planning step</b>	<b>Activity</b>	<b>Interviewee</b>
	Challenge with critical questions.	6
	Incorporate lessons learned.	4, 5
	Design the business model.	4
	Structure bottom-up reporting to senior management.	3
	Consider partnerships in strategy design.	1
Strategy Evaluation	Scenario creation.	3, 4, 6
	Assess the feasibility of a strategy in the market.	2
	Perform risk analysis.	1, 3, 4
	Estimate risk probabilities of goal attainment and margins.	6
	Conduct peer review of proposed strategies.	3
	Scenario evaluation.	4
Strategy Implementation	Manage stakeholders.	1, 2, 3, 4
	Define roles and responsibilities.	1, 3
	Adjust organizational model and governance.	3
	Plan and allocate resources and capacity.	2, 3, 5, 6
	Allocate investments/budgets.	2, 5
	Create shared ownership.	5
	Coach and train teams; facilitate change.	3, 4
	Refine the business model to support execution.	6
Control & Monitor	Define and measure KPIs.	1, 2, 3, 4, 5, 6
	Monitor resources, utilization and deadlines.	1, 3
	Continuously evaluate threats and opportunities.	2
	Run management reviews (QBRs, governance meetings).	2
	Produce evaluation reports.	1, 2, 5
	Detect trend deviations and product risks; take corrective actions.	3, 5
	Adjust strategy based on new trends and customer needs.	3

### **4.1.3 Ranking of strategic planning steps**

All interviewees were asked to rank the top three strategic planning steps based on the perceived usefulness of generative AI for the activities within each step. This prioritization narrows the scope of the broader strategic planning model and informs the focus of the subsequent

quantitative phase, allowing for a more in-depth examination of the most relevant steps. The 1<sup>st</sup>, 2<sup>nd</sup> and 3<sup>rd</sup> choices provided by each participant are presented in Table 4.3.

# Interviewee	1st (3 pts)	2nd (2 pts)	3rd (1 pt)
Interviewee 1	Strategy implementation	Strategy evaluation	Strategy control & monitor
Interviewee 2	Objective setting	Environmental analysis	Strategy evaluation
Interviewee 3	Environmental analysis	Strategy evaluation	Strategy control & monitor
Interviewee 4	Environmental analysis	Strategy evaluation	Strategy formulation
Interviewee 5	Strategy control & monitor	Environmental analysis	Objective setting
Interviewee 6	Environmental analysis	Strategy evaluation	Strategy implementation

Table 4.3: Top-3 rankings of strategic planning steps per interviewee.

In terms of frequency of appearance in participants' top-three ranking shown in Table 4.3, the following observations can be made:

- **Objective setting** was mentioned 2 times (once as 1st choice, once as 3rd choice).
- **Environmental analysis** appeared 5 times (three 2nd-place rankings and two 3rd-place rankings), making it one of the most frequently selected steps.
- **Strategy formulation** was selected once (as 3rd choice).
- **Strategy evaluation** was mentioned 5 times (four 2nd-place rankings and one 3rd-place ranking), indicating a high perceived usefulness of generative AI in this step.
- **Strategy implementation** was selected twice (once as 1st choice and once as 3rd choice).
- **Strategy control & monitor** appeared 3 times (once in each ranking position), suggesting moderate relevance.

Overall, *environmental analysis* and *strategy evaluation* are the most frequently cited steps, each appearing in five out of six participants' top-three rankings. *Strategy control & monitor* follows with three mentions, while the remaining steps were cited less often. These results suggest that, based purely on frequency, participants most consistently associate generative AI with usefulness in *environmental analysis* and *strategy evaluation*.

To identify the most relevant strategic planning steps for further analysis, the interviewees' rankings from Table 4.3 were converted into a quantitative format using the Modified Borda Count (MBC) method. In this scoring approach, 1st-place rankings receive 3 points, 2nd-place rankings 2 points and 3rd-place rankings 1 point. This transformation allows for a data-driven prioritization of steps and the resulting scores are shown in Table 4.4.

# Interviewee	OS	EA	SF	SE	SI	SCM
Interviewee 1	0	0	0	2	3	1
Interviewee 2	3	2	0	1	0	0
Interviewee 3	0	3	0	2	0	1
Interviewee 4	0	3	1	2	0	0
Interviewee 5	1	2	0	0	0	3
Interviewee 6	0	3	0	2	0	0
<b>Total points</b>	<b>4</b>	<b>13</b>	<b>1</b>	<b>9</b>	<b>3</b>	<b>5</b>

Table 4.4: Converted scores using Modified Borda Count.

As shown in Table 4.4, *environmental analysis* received the highest total score, with 13 points, indicating a strong perceived usefulness of generative AI in this step. This score is notably higher than those of the other steps. The second-highest score was obtained by *strategy evaluation*, with a total of 9 points. The remaining steps received considerably fewer points: 5, 4, 3 and 1 respectively. Based on these data-driven results, the subsequent quantitative phase will focus on *environmental analysis* and *strategy evaluation*, as these were identified as the most relevant steps by the interviewees. Qualitative data from the expert interviews support the high score for Environmental Analysis, with Interviewee 6 saying: "If you don't do this, what are you going to build everything on?", referring to environmental analysis. Accordingly, the strategic planning activities to be used in the questionnaire are presented in Table 4.5.

Environmental Analysis (EA)	Strategy Evaluation (SE)
<ul style="list-style-type: none"> <li>• Conduct SWOT analysis.</li> <li>• Conduct PESTLE analysis.</li> <li>• Conduct Porter's Five Forces analysis.</li> <li>• Perform market analysis.</li> <li>• Perform competitor analysis.</li> <li>• Perform stakeholder analysis.</li> <li>• Conduct organizational maturity / AS-IS analysis.</li> <li>• Brainstorm about client/competitor positioning ("market edge").</li> <li>• Continuously scan technological developments.</li> </ul>	<ul style="list-style-type: none"> <li>• Scenario creation.</li> <li>• Assess the feasibility of a strategy in the market.</li> <li>• Perform risk analysis.</li> <li>• Estimate risk probabilities of goal attainment and margins.</li> <li>• Conduct peer review of proposed strategies.</li> <li>• Scenario evaluation.</li> </ul>

Table 4.5: Activities identified for Environmental Analysis and Strategy Evaluation in the interview phase.

## 4.2 Survey results

This section will present the results of the quantitative phase of this research. Building on the exploratory insights from the qualitative interviews, a survey was conducted to evaluate respondents' current satisfaction and perceived future importance of using generative AI across various strategic planning activities. The resulting gap between these two dimensions highlights where generative AI is likely to offer the most added value. Besides, respondents were also asked several questions regarding prompting. This section further presents the data cleaning, demographics and quantitative results both overall and on activity level.

As outlined in the methodology, data cleaning is essential to ensure the reliability and robustness of the dataset. A total of 266 respondents participated in the survey, representing a response rate of approximately 13.5%. Of these, 173 respondents completed the entire survey. To further refine the dataset, respondents who indicated "Never" in response to the question on the frequency of strategic planning were excluded, leading to the removal of an additional six cases and resulting in 167 usable responses. Finally, two attention checks were included in the survey to ensure respondent attentiveness. A total of 128 participants successfully passed both checks. After applying all cleaning criteria, 128 responses were retained for data analysis.

As respondents had the option to select "Not applicable / Not familiar with" for each activity, this response serves as an indicator of whether a strategic planning activity was relevant to their role or if they were familiar with it. It is therefore important to contextualize how many respondents marked this option, as it affects the effective sample size per activity. As a result, only the responses from consultants who indicated familiarity with a given activity were included in the calculation of satisfaction, importance and gap scores. In other words, "Not applicable / Not familiar with" responses were excluded from the statistical analysis per activity, ensuring that the reported results reflect only those participants with relevant experience for each activity. Figure 4.1 illustrates the number of respondents who selected this option across all strategic planning activities, with after each activity: [% of consultants familiar/applicable with / % of consultants not familiar/applicable with]. Interestingly, a relatively high number of consultants indicated unfamiliarity with well-known frameworks such as SWOT, PESTLE and Porter's Five Forces. This stands in contrast to the literature, where these tools are consistently highlighted as foundational to environmental analysis [7, 64, 2]. This finding may suggest a potential divergence between theory and practice: while academia continues to emphasize frameworks like SWOT and PESTLE in their textbook form, consultants may engage with them under alternative names, in simplified format, embedded within client-specific methodologies or that these tools are less frequently employed in daily consulting practice than expected. In contrast, fewer respondents marked "Not applicable / Not familiar with" for market analysis and risk analysis (both with only 29 responses), suggesting broader relevance across roles. Notably, the activity "Brainstorming with client" had the fewest exclusions ( $n = 14$ ), implying that it is the most universally recognized and practiced activity among the sample.

Number of Respondents Marking 'Not Familiar/Applicable' per Strategic Activity

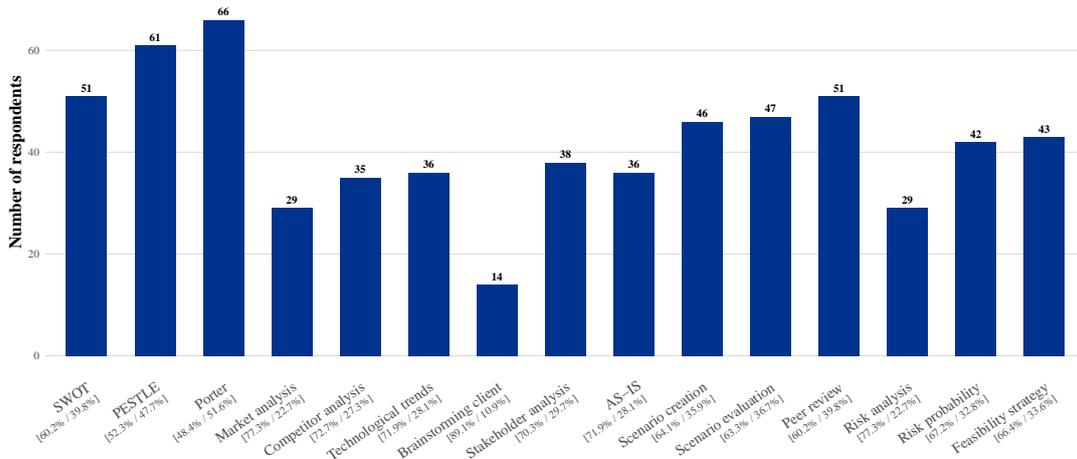


Figure 4.1: Number of respondents indicating 'Not applicable / Not familiar with' per strategic activity.

#### 4.2.1 Demographics

The survey consists of four demographic questions, age, seniority (years of experience), specialization and job title. This section also presents four attitude items regarding generative AI: perceived usefulness, ease of use, perceived risk and future importance. In conclusion, the distribution of the AI usage by consultants and the frequency of engagement in strategic planning activities are presented.

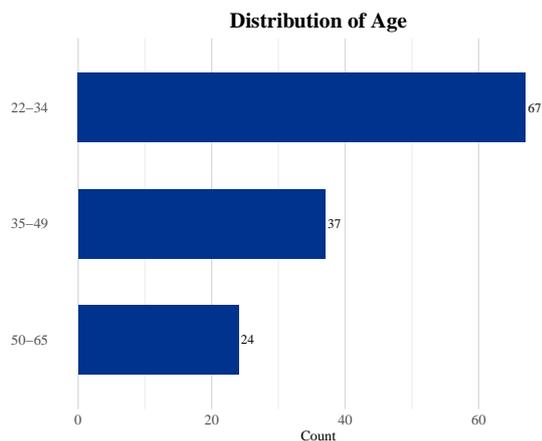


Figure 4.2: Age distribution

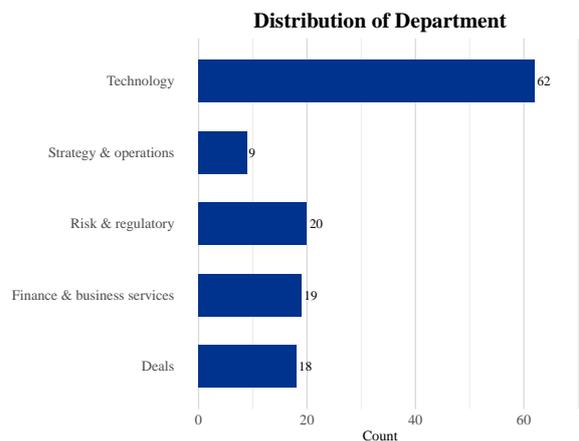


Figure 4.3: Department distribution

The distribution of age among consultants is shown in Figure 4.2. A majority of 67 respondents (52%) are between 22 and 34 years old. This is followed by 37 consultants (29%) aged 35–49 and 24 consultants (19%) aged 50–65. These numbers indicate that most consultants of the population are younger than 50, with over half being under 35.

Figure 4.3 displays the departments in which the consultants are active. The largest share of respondents, 62 consultants (48%), work in the Technology department. The remainder are distributed across Strategy & Operations (9 consultants, 7%), Risk & Regulatory (20 consultants, 16%), Finance & Business Services (19 consultants, 15%) and Deals (18 consultants, 14%). The high proportion from the Technology department is explained by the researcher’s affiliation with this group which led to targeted distribution within this department.

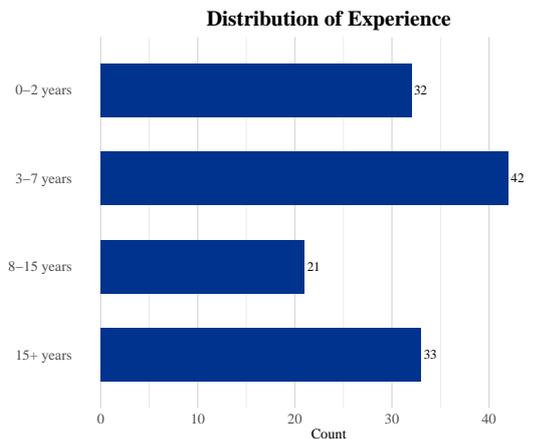


Figure 4.4: Experience level distribution

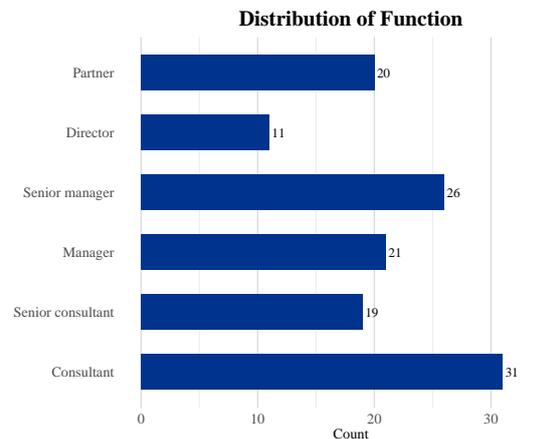


Figure 4.5: Function types distribution

Figure 4.4 illustrates the respondents’ professional experience in consulting and strategic planning. A total of 32 respondents (25%) have 0–2 years of experience, 42 (30%) have 3–7 years and 21 (15%) have 8–15 years. The remaining 33 respondents (25%) possess over 15 years of experience, representing the most senior group. This variation demonstrates a wide range of seniority levels across the sample.

Finally, Figure 4.5 shows the functional roles of the respondents. There are 31 consultants (24%), 19 senior consultants (15%), 21 managers (16%), 26 senior managers (20%), 11 directors (9%) and 20 partners (16%). The sample includes respondents across all seniority levels, with both junior and senior roles represented. Partners make up a relatively large share of the respondents compared to senior consultants. This is remarkable given the typical pyramid structure of Big Four firms, where the number of professionals decreases sharply at higher ranks. In such structures, the number of senior consultants usually far exceeds that of partners, making their equal representation in this sample noteworthy.

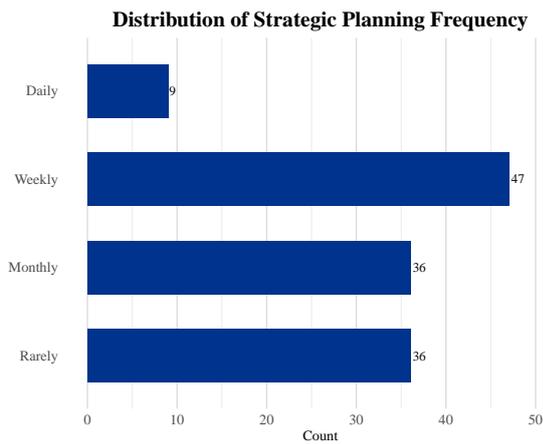


Figure 4.6: Distribution of AI usage

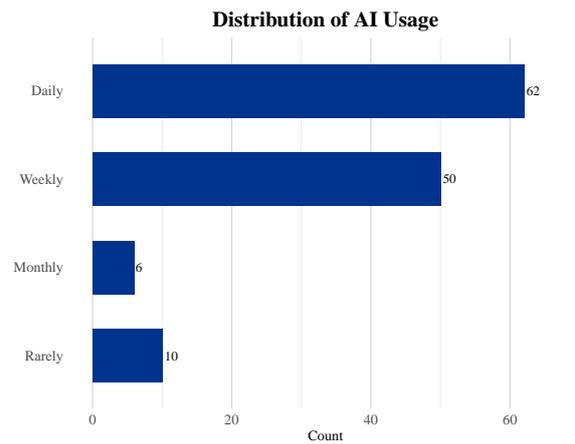


Figure 4.7: Strategic planning frequency

Figure 4.6 shows the frequency of generative AI usage among respondents. A total of 62 respondents (48%) indicated daily use, followed by 50 respondents (39%) who reported weekly use. A smaller number reported using generative AI monthly (6 respondents, 5%) or rarely (10 respondents, 8%). These figures indicate that the majority of respondents engage with generative AI on a regular basis.

Figure 4.7 presents the frequency of engagement in strategic planning activities. Weekly participation was reported most frequently (47 respondents, 37%), followed by both monthly and rarely, each with 36 respondents (28%). Daily involvement was least common (9 respondents, 7%). These results show that while strategic planning is a regular aspect of work for many respondents, it is not typically performed on a daily basis.

### 4.2.2 Attitudes toward generative AI

All attitude items were measured using a 7-point Likert scale, ranging from 1 (not at all) to 7 (to an extremely large extent). Figure 4.8, Figure 4.9, Figure 4.10 and Figure 4.11 display the average scores for each attitude category, segmented by experience, department, age group and job function, respectively. These figures report responses across four attitude dimensions related to generative AI: ease of use, future importance, perceived risk and perceived usefulness. Each figure includes the corresponding sample size for each subgroup, denoted as [n], allowing for comparison between different respondent profiles.

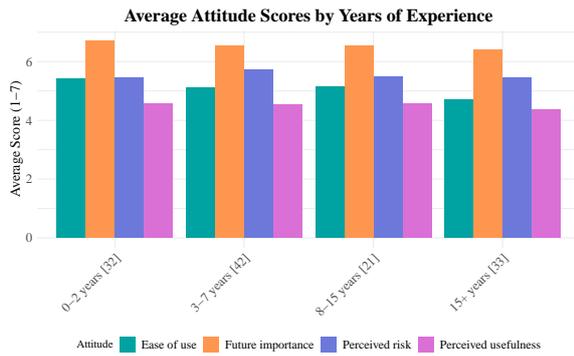


Figure 4.8: Attitude scores by Experience

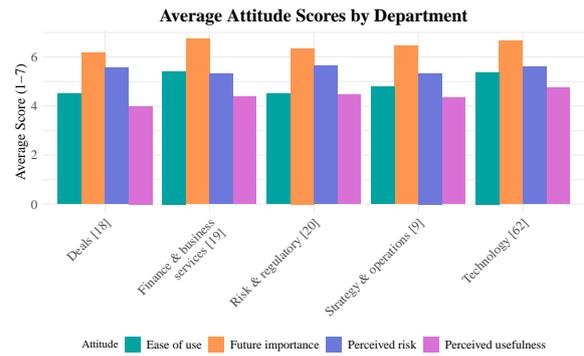


Figure 4.9: Attitude scores by department

Figure 4.8 shows a consistent pattern across experience levels, with all groups assigning high scores to the future importance of generative AI in their professional work. At the same time, the current perceived usefulness was rated notably lower across all groups, creating a visible gap between anticipate future value and current satisfaction. A modest downward trend is observable in ease of use as experience increases. This decline appears more pronounced among those with more than 15 years of experience.

In Figure 4.9, future importance again consistently exceeds a score of 6 on the 7-point Likert scale across all departments. The Technology department reported the highest average perceived usefulness. While ease of use remained relatively consistent, slight differences emerged; respondents from the Technology and Finance & Business services departments gave marginally higher ratings than other groups. Perceived risk was highest in the Risk & Regulatory department and lowest in Strategy & Operations, indicating variation in risk perceptions between domains.

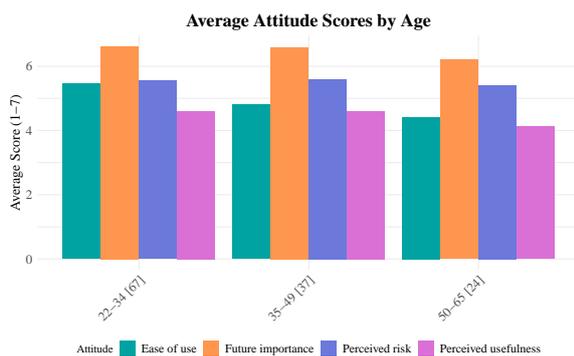


Figure 4.10: Attitude scores by age

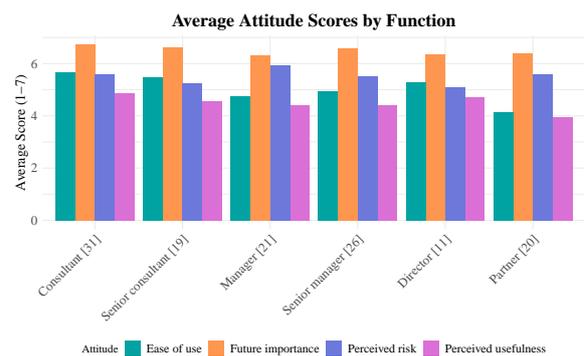


Figure 4.11: Attitude scores by function

A similar pattern is observed in Figure 4.10 and Figure 4.11. Future importance remains consistently high across all age and function groups, while perceived usefulness shows a more moderate trend and has the lowest average across the four attitude items. A visible decline in ease of use emerges after age 35, marking a clear difference between younger and older

respondents. Perceived risk is relatively stable across age groups but fluctuates more across functions. Two modest drops are seen, between consultant and manager and again from manager to partner, before increasing at partner level. Figure 4.11 further illustrates that ease of use tends to be rated highest by consultants and senior consultants, before gradually declining, with a notably sharp drop from director to partner.

### 4.2.3 Current state of prompting and enablement

This section delves deeper into prompting, specifically examining whether respondents have received training on the use of generative AI and how they rated this training. If they would benefit from tailored prompting strategies. Various prompting techniques are asked if they know them. Lastly, respondents were asked how they craft their prompts.

Respondents were asked whether they had received any training related to generative AI. As this survey item allowed multiple selections, the total number of responses in Figure 4.12 exceeds the number of individual participants. Each response category is shown on the vertical axis, with the corresponding number of respondents indicated in square brackets (e.g., [n = 50]) and the percentage displayed at the end of each bar. The largest share (39.1%) reported receiving internal training. In addition, 21.1% indicated that they had received both internal and external training. A smaller group (8.6%) had participated in external training only, while 31.2% had not received any training at the time of the survey.

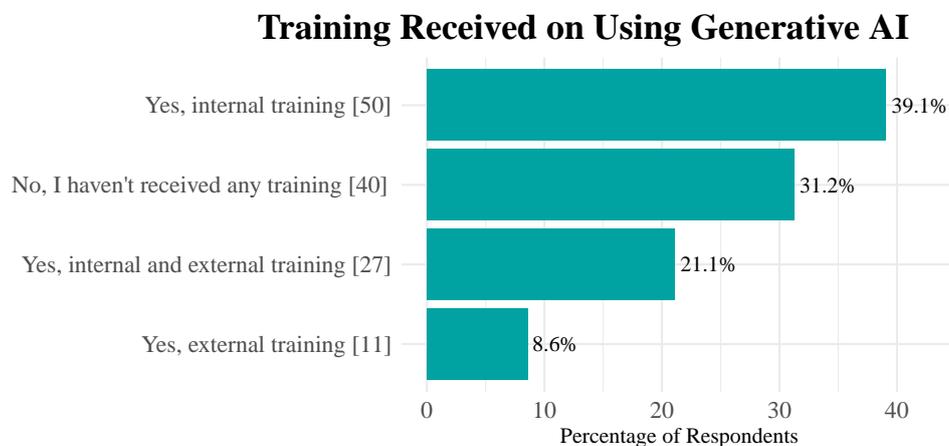


Figure 4.12: Share of respondent have received internal or external training on the use of Generative AI.

Figure 4.12 shows that a majority of consultants have already received training on the use of generative AI, with internal training being the most common. Figure 4.13 presents respondents' perceived usefulness of the most recent AI training they received. The responses vary considerably: while nearly half of the consultants rated the training as useful or very useful, a substantial portion reported only moderate or limited usefulness. This distribution suggests heterogeneous experiences with existing training programs.

In contrast, Figure 4.14 shows a strong positive consensus regarding the perceived value of tailored prompting strategies in the context of strategic planning. The majority of respondents agreed with this statement, indicating widespread interest in more targeted prompting support among the consultants surveyed.

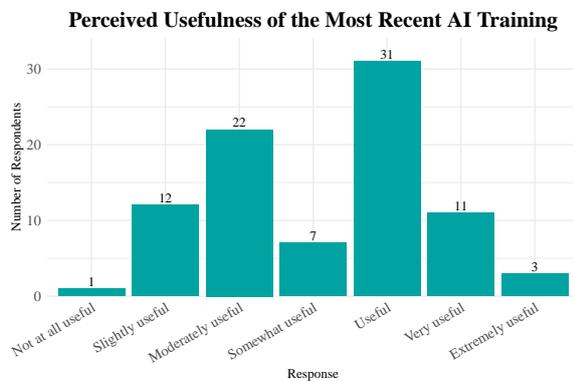


Figure 4.13: Usefulness AI training

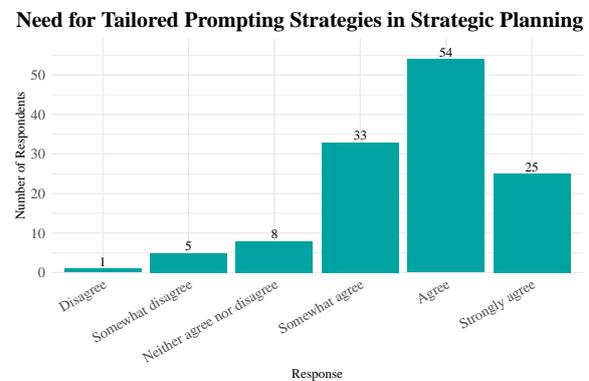


Figure 4.14: Need for prompting strategies

Figure 4.15 and Figure 4.16 are both multiselect survey items, therefore having the same structure as Figure 4.12. In terms of support needs for effective generative AI use (Figure 4.15), the most frequently selected enablement resource was hands-on training in writing effective prompts, accounting for 23.3% of total mentions. This was followed by prompting templates or examples and access to better tools or platforms (both 19.6%). Clear internal guidelines or best practices (15.2%) and time or space to experiment with AI in client work (13.4%) also received notable mention. A relatively small share of mentions referred to a community to share knowledge (8.9%). Notably, the prominence of workshops and example-based prompting resources suggests a relatively strong preference for practical, guided forms of enablement. Compared to more passive or exploratory options, such as documentation, experimentation time or community forums, support formats that involve direct instruction or demonstration were selected more frequently.

Figure 4.16 presents how consultants currently engage in prompting. The most common strategy, selected in 45.6% of total mentions, was creating prompts from scratch. Using whatever comes to mind in the moment followed with 30.1%, while not using any specific prompts accounted for 11.5%. Less frequently mentioned were using prompts found online or from others (8.8%) and having prompts tailored to specific strategic tasks (4%).

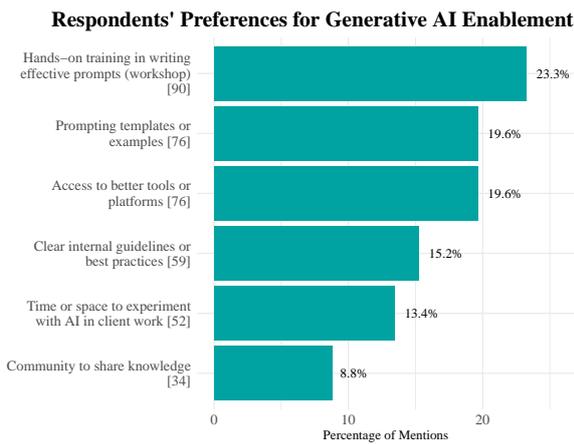


Figure 4.15: Consultant’s support needs for Generative AI use

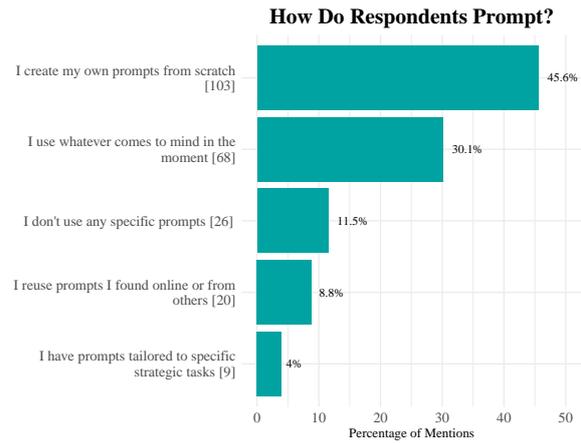


Figure 4.16: Current prompting practices among consultants

#### 4.2.4 Gap Analysis of current satisfaction & future importance

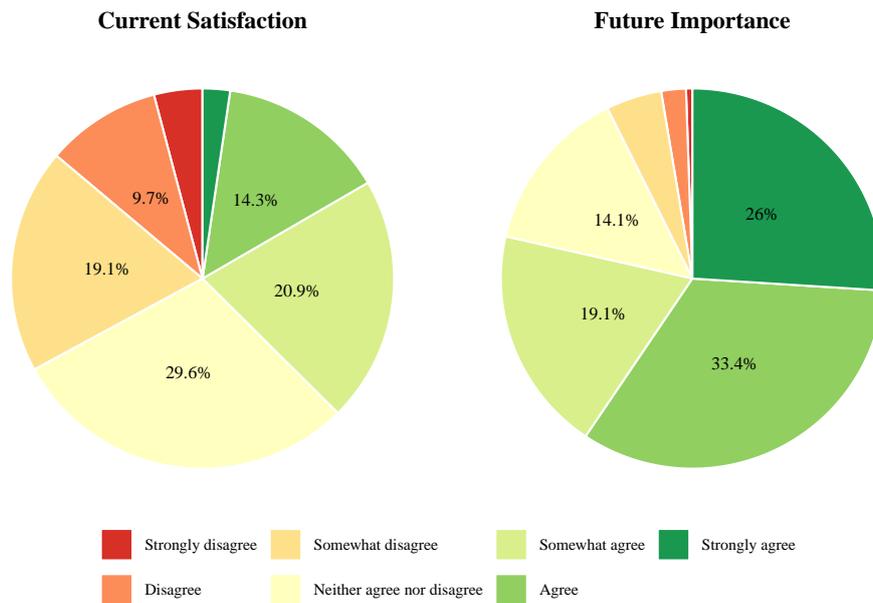
This section presents the main quantitative findings of the study, based on the paired survey items measuring consultants’ current satisfaction and perceived future importance of generative AI for various strategic planning activities. For each activity, mean satisfaction, mean importance and the resulting gap scores are reported alongside their respective standard deviations (SD) and sample sizes. These results provide a detailed overview of where perceived potential for generative AI application is highest or lowest across the surveyed activities. In all tables and figures, gap denotes the mean of respondent-level paired difference.

First, the results of the paired-item questions across all respondents and strategic planning activities are presented in the two pie charts in Figure 4.17. For clarity, response options representing less than 5% of total responses are excluded from the visualizations.

As shown in Figure 4.17, current satisfaction is most commonly rated as "Neither agree nor disagree" (29.6%), followed by "Somewhat agree" (20.9%) and "Somewhat disagree" (19.1%). This indicates that most consultants hold a neutral or mildly opinionated stance regarding their current satisfaction with generative AI. The options "Agree" (14.3%) and "Disagree" (9.7%) follow, while the two extremes "Strongly agree" and "Strongly disagree" are the least selected, suggesting a relatively normal and balanced distribution.

In contrast, when considering future importance, the responses show a clear shift toward more positive expectations. The majority of respondents selected "Agree" (33.4%) or "Strongly agree" (26.0%), together accounting for nearly 60% of all responses. These are followed by "Somewhat agree" (19.1%) and "Neither agree nor disagree" (14.1%). The remaining disagreement-oriented categories ("Somewhat disagree", "Disagree" and "Strongly disagree") each accounted for less than 5% of responses. This distribution shows a clear skew toward agreement, with most consultants rating generative AI as important for future strategic planning.

## Distribution of Likert Scores



*Note: Scores < 5% are not labeled in the chart for readability.*

Figure 4.17: Distribution of current satisfaction and future importance scores among all consultants.

Second, the results of the paired-item questions per activity will be shown in Table 4.6. The table is sorted in descending order based on the gap values. It provides an overview of key quantitative results, listing each strategic planning activity alongside its corresponding strategic planning step. For each activity, the mean satisfaction, importance and gap scores are reported, together with their respective SDs. The final columns show the number of valid responses per activity, which varies across items as discussed in Section 4.2. To support interpretation, this table is also visualized in Figure 4.18, where the satisfaction, gap and importance scores are depicted graphically alongside their underlying satisfaction and importance values.

Activity	Step	M_Sat	SD_Sat	M_Imp	SD_Imp	Gap	SD_Gap	n
Competitor analysis	Environmental Analysis	4.01	1.31	5.80	1.05	1.78	1.33	93
Risk analysis	Strategy Evaluation	4.04	1.24	5.73	1.06	1.69	1.26	99
Risk probability	Strategy Evaluation	3.57	1.44	5.26	1.41	1.69	1.54	86
Scenario creation	Strategy Evaluation	3.82	1.43	5.43	1.40	1.61	1.46	82
Technological trends	Environmental Analysis	4.10	1.32	5.70	1.24	1.60	1.41	92
Scenario evaluation	Strategy Evaluation	3.90	1.32	5.49	1.25	1.59	1.40	81
Market analysis	Environmental Analysis	4.44	1.33	5.96	1.07	1.52	1.44	99
Feasibility strategy	Strategy Evaluation	3.52	1.55	4.94	1.56	1.42	1.48	85
Stakeholder analysis	Environmental Analysis	4.06	1.48	5.44	1.32	1.39	1.51	90
SWOT	Environmental Analysis	4.47	1.15	5.81	1.09	1.34	1.15	77
PESTLE	Environmental Analysis	4.55	1.16	5.88	0.93	1.33	1.26	67
Porter	Environmental Analysis	4.55	1.18	5.87	1.00	1.32	1.23	62
Peer review	Strategy Evaluation	3.83	1.57	5.16	1.48	1.32	1.41	77
AS-IS	Environmental Analysis	3.77	1.34	5.07	1.54	1.29	1.44	92
Brainstorming client	Environmental Analysis	4.32	1.42	5.55	1.36	1.23	1.37	114

Table 4.6: Gap analysis of generative AI usefulness across strategic planning activities.  $M\_Sat$  and  $M\_Imp$  indicate mean satisfaction and importance;  $Gap$  is the mean of respondent-level paired difference (Imp - Sat);  $SD\_Sat$ ,  $SD\_Imp$ , and  $SD\_Gap$  report standard deviations;  $n$  is the number of responses.

As shown in Table 4.6, there are nine environmental analysis activities and six strategy evaluation activities. The SD reflects how much responses vary across consultants. A higher SD indicates less agreement, pointing to greater ambiguity or inconsistency in how the usefulness of generative AI is perceived for that activity.

On a 7-point Likert scale, the maximum SD is 3, which occurs when responses are evenly split between the two extremes: 1 and 7. This can be illustrated by the following calculation:

The mean can be calculated via:

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i \Rightarrow \mu = \frac{1+7}{2} = 4$$

The variance is as follows:

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 \Rightarrow \sigma^2 = \frac{(1-4)^2 + (7-4)^2}{2} = \frac{9+9}{2} = 9$$

And the maximum SD is then:

$$\sigma = \sqrt{\sigma^2} \Rightarrow \sigma = \sqrt{9} = \boxed{3}$$

In practice, SD values above 1.5 already signal substantial disagreement. Table 4.7 highlights the three activities with the highest and lowest average SD across satisfaction, importance and gap scores. The top three indicate lower consensus and more varied perceptions of generative AI's usefulness. The bottom three show high agreement, suggesting a shared understanding

among consultants of how generative AI supports those specific activities. Interestingly, Porter, SWOT and PESTLE all structured analytical frameworks show the lowest SDs, indicating high agreement on the perceived usefulness of generative AI for these activities. Conversely, determining the feasibility of a strategy, conducting peer review and risk probability are associated with the highest SDs, suggesting greater ambiguity or inconsistency among respondents.

Group	Activity	SD_Sat	SD_Imp	SD_Gap	Mean_SD
High variability	Feasibility strategy	1.55	1.56	1.48	1.53
High variability	Peer review	1.57	1.48	1.41	1.49
High variability	Risk probability	1.44	1.41	1.54	1.46
Low variability	Porter	1.18	1.00	1.23	1.14
Low variability	SWOT	1.15	1.09	1.15	1.13
Low variability	PESTLE	1.16	0.93	1.26	1.12

Table 4.7: Top 3 and Bottom 3 activities by average standard deviation across satisfaction, importance, and gap

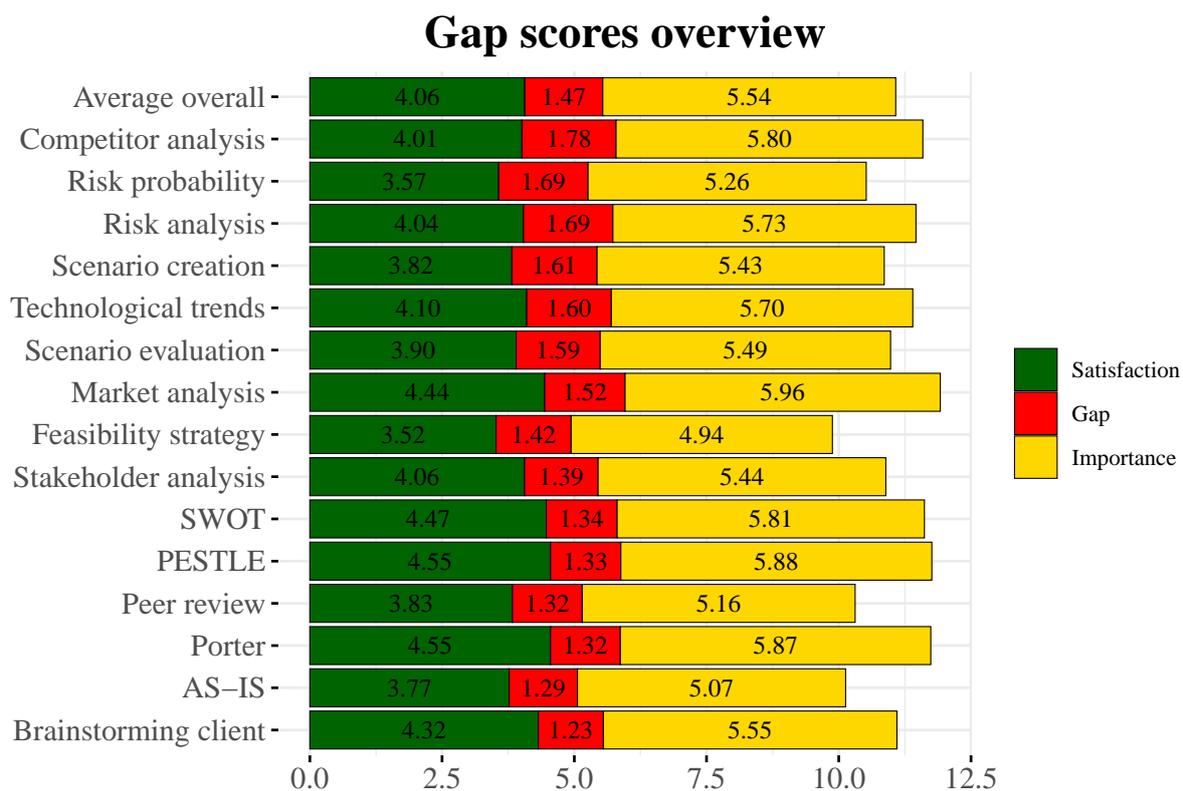


Figure 4.18: Gap analysis illustrating the gap scores, including satisfaction and importance for all strategic planning activities.

Figure 4.17 presented the overall distribution of responses, while Table 4.6 detailed satisfaction, importance and gap scores per activity. Figure 4.18 now visualizes these results, showing each activity's satisfaction, importance and resulting gap score, ranked by gap magnitude. This figure complements the numeric table by making relative differences more immediately apparent.

Satisfaction scores across all strategic planning activities range from 3.52 to 4.55, while importance scores range from 4.94 to 5.96. Corresponding gap values fall between 1.23 and 1.78. This consistent difference of at least one Likert scale point between satisfaction and importance suggests that for every activity, consultants perceive a notable degree of unrealized potential in current generative AI support.

The three largest gaps are observed in competitor analysis (1.78), risk probability (1.69) and risk analysis (1.69). These activities combine relatively low satisfaction scores, particularly in the case of risk probability (3.57), with importance scores ranging around 5.80 and 5.26 for risk probability. The magnitude of these gaps signals a clear view among respondents that current generative AI support in these activities lags behind its perceived future relevance.

By contrast, activities such as SWOT analysis, PESTLE and Porter exhibit much smaller gaps, ranging from 1.32 to 1.34. These activities are also among those with the highest satisfaction and importance ratings. Indicating that these activities have a relatively close alignment between existing generative AI support and future expectations in these well-established structured analytical frameworks.

The smallest gap is observed in brainstorming over a client (1.23), followed by AS-IS analysis (1.29). Notably, the former is associated with a higher satisfaction score (4.32 vs. 3.77) and a higher importance score (5.55 vs. 5.07) compared to the latter.

Several activities exhibit both high gap scores and high importance scores. These include competitor analysis, risk analysis, technological trends and market analysis. These activities all score above 5.70 importance and have gap values exceeding 1.50. Satisfaction for these activities remain low around 4.10, except for market analysis (4.44). Market analysis has the highest future importance score (5.96) as with relatively high satisfaction and gap values. Together, these patterns suggest that these activities may offer the most pronounced opportunities for generative AI support for strategic planning activities.

Lastly, it is noteworthy to mention that assessing the feasibility of a strategy has both the lowest satisfaction (3.52) and importance (4.94) score among all activities, while its gap value remains close to the overall average.

Figure 4.19 presents the gap scores grouped by strategic planning step, comparing activities categorized under Environmental Analysis (EA) and Strategy Evaluation (SE). A clear difference emerges between the two groups in terms of satisfaction, importance and gap scores. The average satisfaction score for EA activities is 4.25, whereas SE activities average 3.78, nearly half a Likert scale point lower. This suggests that consultants are generally more satisfied with the current generative AI support for EA tasks than for SE tasks.

## Gap Scores Overview by Strategic Planning Step

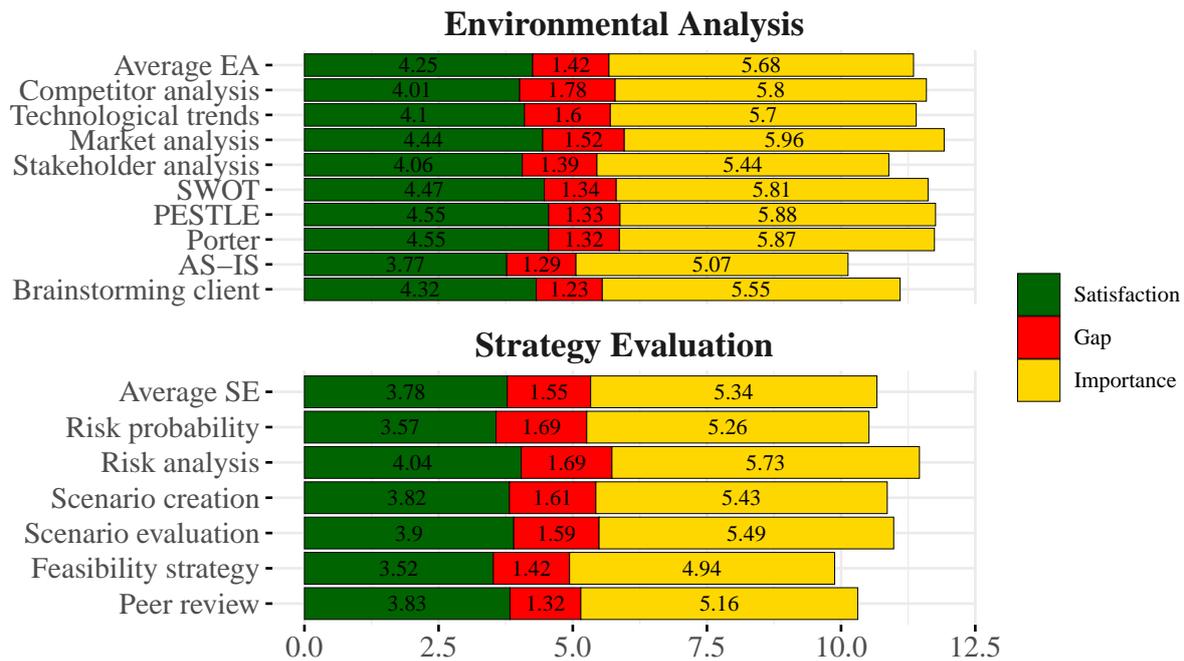


Figure 4.19: Gap analysis illustrating the gap scores, including satisfaction and importance for all strategic planning activities by each corresponding step.

Similarly, the perceived future importance of generative AI is higher for EA (5.68) compared to SE (5.34). While both steps are rated above the midpoint of the scale, the higher average for EA has an indication that consultants assign relatively greater value to generative AI's role in environmental scanning, market insights or external analyses.

In terms of the gap between satisfaction and importance, SE again scores less favorably, with an average gap of 1.55 compared to 1.42 for EA. This indicates that SE tasks are more frequently perceived as areas where current generative AI support falls short of its expected usefulness, relative to EA tasks.

Taken together, these results indicate a consistent pattern: consultants report higher satisfaction and importance for generative AI within EA activities, while SE tasks show lower satisfaction and wider perception gaps.

### 4.2.5 In-depth gap analysis based on demographics

The demographics can be used to provide in-depth analysis into the gap analysis. To avoid overwhelming the main narrative with detailed subgroup visualization, only one functional breakdown is shown in the Results section in Figure 4.20. This figure illustrates how respondent demographics can enrich the interpretation of gap patterns across strategic planning activi-

ties. Additional subgroup heatmaps (e.g., by seniority, age and department) are included in Appendix E (See Figures E.1, E.2 and E.3) for references and completeness.

Satisfaction scores show a declining trend with age. Respondents aged 25–34 report the highest average satisfaction ( $M = 4.20$ ), while those aged 35–50 score slightly lower ( $M = 4.12$ ) and respondents aged 50–65 report the lowest satisfaction overall ( $M = 3.53$ ). This pattern is consistent across nearly all activities and is particularly pronounced in items such as scenario creation, scenario evaluation and feasibility strategy.

The gap values increase as consultants become older. Notably, the jump from the 22–34 group to 50–65 is from 1.12 to 2.03, indicating a clear difference in gap between age groups.

Importance scores are generally stable across age groups, though slightly elevated among respondents aged 35–50. For example, for the activity SWOT analysis, the importance score is 6.27 in the 35–50 group compared to 5.55 in the 25–34 group and 5.54 in the 50–65 group. This results in the largest average gap value also occurring in the 35–50 group ( $M = 1.64$ ).

Some activities exhibit particularly large gaps for specific age groups. For instance, scenario evaluation has a gap of 2.33 in the 50–65 group and scenario creation shows the highest gap overall in the 50–65 group (2.67). In contrast, the smallest average gap is observed in the 25–34 group (0.82), which is also the group with the highest satisfaction levels.

Figures E.1 and E.2 display similar patterns as seen in Figure 4.20. As shown in Figure E.1, gap scores increase progressively with function, ranging from an average of 0.96 for consultants to 2.15 for partners. This pattern is mirrored by declining satisfaction scores, with a notable drop among senior roles. Directors show slightly higher satisfaction than expected within this trend, forming a mild outlier. Importance scores remain relatively stable across functions, hovering around 5.60, though consultants again stand out with a lower average importance score of 5.21.

Interestingly, directors report exceptionally low gap values for structured analytical frameworks like SWOT (0.80), PESTLE (0.50), Porter (0.75) and Market analysis (0.57). These activities are also associated with high satisfaction scores, ranging from 4.80 to 5.86, suggesting strong alignment between expectations and current generative AI capabilities in these areas for this group.

Turning to Figure E.2, a similar trend is evident. Gap scores rise steadily with experience level, from 1.08 among consultants with 0–2 years of experience to 2.04 for those with 15+ years. Satisfaction scores decline in parallel, particularly dropping after the 15-year mark. Importance scores remain relatively stable but are slightly elevated in mid-career respondents (3–7 years and 8–15 years), structured analytical frameworks.

These patterns collectively suggest that as consultants become more experienced or assume more senior roles, they tend to perceive a wider gap between the current support generative AI offers and what they consider important for strategic planning tasks, while also expressing lower satisfaction.

Figure E.3 shows the detailed gap, satisfaction and importance values for the departments. The Strategy & Operations department accounts for the highest average gap value (2.07), while Deals has the lowest average gap value (1.16). The Strategy & Operations department shows the

## Activity Scores by Age

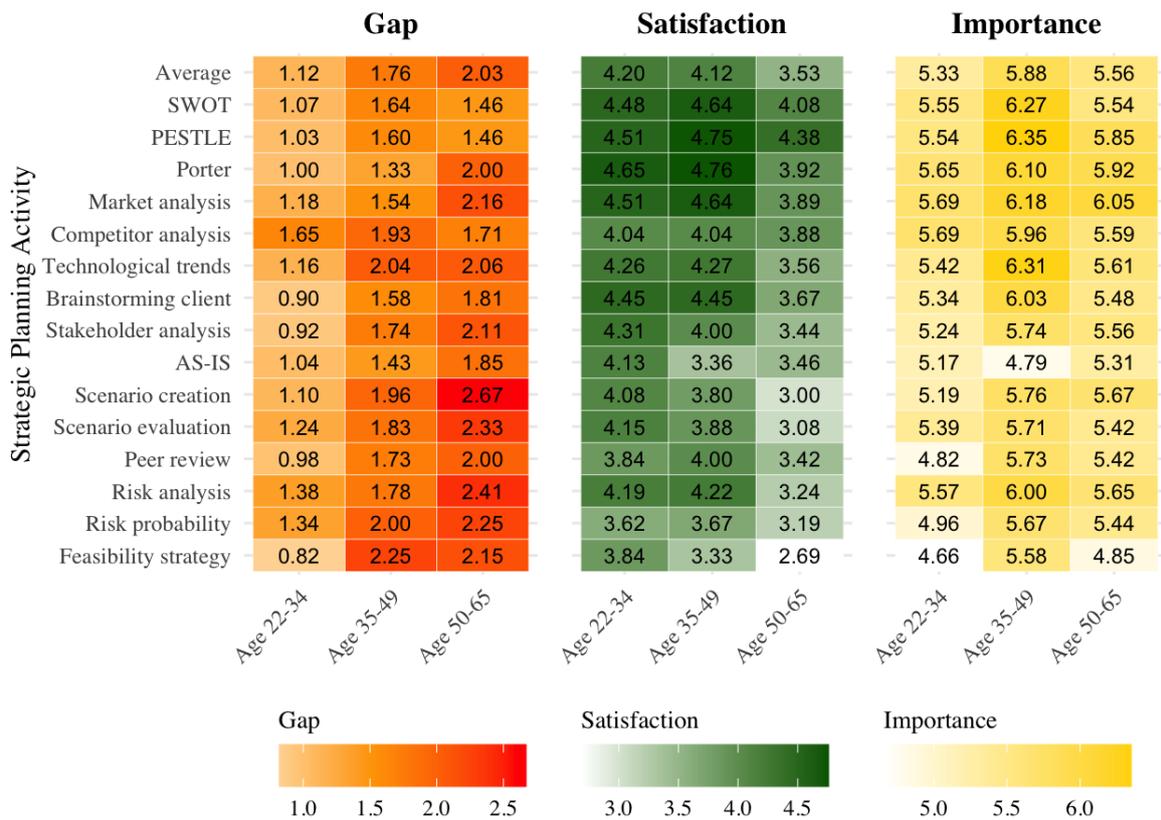


Figure 4.20: Heatmap of gap, satisfaction, and importance scores for each strategic planning activity, broken down by age group. Values represent mean scores per age. Darker shades indicate more extreme values within each scale.

lowest satisfaction scores among the figure. Deals has some of the lowest importance scores, having as low as 4.00 for AS-IS analysis.

# Chapter 5 Discussion

This chapter draws on the results presented in Chapter 4 and aims to interpret the findings, identify the key insights and place them in the context of the main research question. The discussion begins by examining the benefits and challenges of using generative AI in strategic planning, thereby providing a direct answer to the central research question: *what are the benefits and challenges of using Generative AI to support strategic planning activities in management consulting?* In addition to addressing the research question, the chapter also considers insights that emerged beyond its direct scope. These findings will also be discussed. By incorporating these points, the discussion not only answers the main question but also highlights areas of broader relevance for both academic and practical contexts. Finally, the chapter reflects on the limitations of the study and outlines opportunities for future research.

## 5.1 Generative AI's role in strategic planning

The research question is addressed in two parts. First, the benefits regarding the usage of generative AI to support strategic planning activities as perceived by the management consultants will be discussed. In this research, benefits refer to those strategic planning activities where consultants reported that generative AI already provides meaningful support. This is reflected in high satisfaction scores, high importance and relatively low gap values between current satisfaction future importance. Second, the challenges are examined. These refer to areas where generative AI is not performing well, but is expected to in the future, typically characterized by low satisfaction, high importance and larger gap values.<sup>1</sup>

### 5.1.1 Benefits

There is a clear shift from current satisfaction toward future importance across all activities. Consultants across all demographics share the belief that generative AI will become increasingly relevant for strategic planning, with the average future importance ( $M = 5.54$ ) rated considerably higher than current satisfaction ( $M = 4.06$ ). This consistent pattern indicates that management consultants in are generally optimistic about the future potential of generative AI, even if current applications are not yet fully optimized. Three key findings will now be distinguished as benefits: (1) structured analytical frameworks as success stories, (2) promising emergent activities and (3) generative AI as a sparring partner.

---

<sup>1</sup>Although “benefits” and “challenges” were not formally operationalized as categories in the methodology, they are interpreted here on the basis of survey outcomes (satisfaction, importance, gap analysis) in order to provide a structured answer to the research question.

## Structured analytical frameworks as success stories

Structured analytical frameworks, such as SWOT, PESTLE and Porter's Five Forces emerge as the clearest success stories in the use of generative AI for strategic planning. Consultants reported some of the highest satisfaction scores in the survey (all around  $M = 4.50$ ) alongside high importance ratings (all above  $M = 5.80$ ), while producing among the lowest gap values ( $M = 1.32 - 1.34$ ). These activities also recorded the smallest standard deviations of all items, signaling strong agreement across age groups, seniority levels and departments. In other words, consultants broadly concur that generative AI already supports these frameworks effectively, positioning them as established use cases rather than experimental ones. Although interestingly, only half of the respondents reported using or being familiar with this set of activities, the literature positions them as a core element of strategic planning practice [71, 48]. Yet, Mintzberg famously critiqued such reliance on formal tools, arguing that real strategy often emerges through adaptive iterative processes [55]. The tension between theory and practice may help explain why only half of the consultants in this study reported usage or familiarity with these frameworks, despite their strong alignment with generative AI.

The reasons for this maturity lie in the characteristics of these tools. SWOT, PESTLE and Porter are highly structured, rule-based frameworks that can be decomposed into predefined categories [7, 2, 64]. This reduces ambiguity and limits the scope of interpretation, which is precisely where LLMs perform most reliably: structured, boxed task. Their input-output logic is simple, repeatable and widely documented. Numerous illustrative and real-world examples are publicly available, meaning that LLMs have likely encountered such patterns extensively during training. This abundance of structured exemplars makes it easier for LLMs to generate coherent and relevant outputs, as they can draw on a broad base of similar cases [1]. A consultant captured this sentiment directly, stating: "I used Copilot for SWOT evaluation. When writing a good prompt and context (without disclosing the client) the results were amazing. After several iterations and refinements the output was great." This illustrates not only how generative AI is already seen as a reliable option for structured analyses, but also once again reiterates prompting remains a critical digital competence.

Other respondents echoed this perception, noting the usefulness of generative AI in quickly populating analytical templates. As one survey participant wrote, "I used Generative AI a[s] source to gain more information on creating a SWOT analysis for a client." Another reflected: "(...) using [LLMs in a] SWOT-analysis kind of activity, where Generative AI helped me (...) for the company. I did find some errors in this though so this highlights the importance of double checking." These comments reinforce the survey's quantitative findings: consultants already view generative AI as a valuable accelerator for structured analytical frameworks, while maintaining awareness of its current limitations. This shows that consultants trust generative AI to accelerate analysis, but not to deliver unverified final outputs.

This finding is also consistent with academic discussions on AI in knowledge-intensive work. Prior research shows that structured, data-driven tasks are more easily automated [67]. Structured analytical frameworks thus represent a natural entry point for generative AI, where

its strength in pattern recognition and templated reasoning can be reliably leveraged. This key finding thus represents the clearest mature use cases (i.e., structured analytical frameworks), where generative AI already delivers consistent and reliable value.

### **Promising emergent activities as generative AI's next frontier**

Another cluster of activities points to what may be described as promising emergent activities. These include competitor analysis, market analysis, risk analysis and technological trend forecasting. Across these items, consultants reported satisfaction scores around the overall mean ( $M = 4.05$ ), with market analysis performing slightly higher and consistently high importance values ( $M > 5.70$ ). The resulting gaps were noticeably larger than for structured frameworks ( $\Delta \approx 0.3$ ), indicating a significant difference between current satisfaction and future expectations. Standard deviations were also somewhat higher, suggesting less consensus across demographics. Together, these patterns highlight a set of activities where consultants are moderately satisfied with generative AI today, but perceive considerable untapped potential for the future, due to its exceptionally high importance scores.

The qualitative data reinforces this interpretation. One consultant observed: "It is a good idea [to use LLMs] for competitor analysis, market analysis (...) if GenAI is able to access that data." Stressing that the LLMs must first gain reliable access to relevant information, which represents a key area for improvement. Another respondent referred to "risk identification, Gen AI is good in providing high-level risks in different areas. Some of which can be overlooked easily." This highlights that risk analysis holds clear potential, but it still has "high-level risks", indicating that LLMs still provide general answers. Others reported using generative AI "on a daily basis for maturity assessments as well as identifying market and technology trends" and described it as "extremely useful and can save a lot of time (...). It is however key that experienced professionals still review (...)." Thus, consultants see early value in speed and breadth, but reliability remains conditional on access to external data and expert oversight.

These quotes illustrate how consultants are already experimenting with AI in these domains, often finding value in speed and idea generation. Yet, unlike the analytical structured frameworks, these activities are less standardized and often highly context-dependent [83, 55]. They involve open-ended synthesis, evaluation of dynamic market signals and room for interpretation, conditions under which LLMs are more likely to produce generic or incomplete outputs. Moreover, the tasks themselves involve retrieving data, synthesizing insights and producing preliminary analysis. While such functions are in principle automatable, consultants consistently stress the need for human oversight, fact-checking and contextual adaptation. This dynamic may explain why satisfaction remains relatively low. Consultants express skepticism, ( $M$  perceived risk = 5.2) despite consistently rating these activities as highly important, reflecting limited trust yet in generative AI for less structured tasks.

Academic literature supports this distinction; studies show that generative AI performs reliably on structured, templated tasks but struggles in domains that require contextual reasoning or creative knowledge [67]. Likewise, research on strategic planning has shown that these

activities are not purely analytical but involve interpretation, human judgment and organizational context [55, 59]. This aligns with the survey's findings that consultants remain cautious when tasks require contextual interpretation rather than template-based reasoning.

There is a clear appetite for generative AI for these promising emergent activities. They are already perceived as somewhat useful, but their true potential lies in the future. Current tools have not yet met expectations and bridging this gap may require both technological improvements and user education, particularly through tailored prompting strategies. Figure 4.14, supports this, showing that the significant majority (> 89%) of consultants believe customized prompting approaches are essential. Providing such training workshops could help these activities from an early stage of adoption toward higher levels of maturity, especially if technological capabilities continue to advance. To sum, these promising emergent activities highlight generative AI's potential: while current performance falls short, consultants remain optimistic about their future value, making them promising yet still immature use cases.

### **Generative AI as a sparring partner for client-focused brainstorming**

Lastly, the activity of brainstorming client needs deserves an honorable mention. With nearly universal adoption among consultants ( $n = 118$ ), this activity received consistently high satisfaction ( $M = 4.32$  and importance scores ( $M = 5.55$ , while recording the lowest gap value ( $M = 1.23$  across all items. In other words, brainstorming represents a "sweet spot" where generative AI is already perceived as both valuable and reliable, with minimal distance between current performance and future expectations. Although the standard deviation was somewhat higher than in the structured analytical frameworks, the consistently positive evaluations indicate that brainstorming has become an established and trusted application of generative AI in consulting practice. The open responses illustrate why. One consultant described the experience as "like having a supercharged brainstorming partner(...) It doesn't replace human judgment, but it definitely makes the process smarter and more dynamic." Another referred to it as "almost like a sparring partner" that supports ideation prior to team discussions. Respondents repeatedly emphasized the creativity boost, for example noting that "it can definitely help while brainstorming, as it is generating ideas, that can be the starting point for new" and even that "I love how it increase creativity and provided broader ideas during brainstorm." These reflections underline why brainstorming has a unique position in the results: it exemplifies how generative AI can reliably accelerate idea generation without undermining human judgment. Rather than replacing consultants, it expands the breadth of options considered, enhances creativity, making it a clear example of generative AI already adding tangible value to strategic planning practice. In this sense, brainstorming stands out as a special case where generative AI is not merely a tool, but a trusted partner in creativity and ideation, while still relying on human judgment to validate and refine ideas.

It is worth noting that nearly all of the activities classified as benefits fall within the domain of Environmental Analysis (EA). This includes both mature frameworks such as SWOT, PESTLE and Porter, as well as promising emergent activities like market and competitor analyses,

technological trend forecasting and brainstorming over clients. EA focuses on assessing internal and external influences through structured tools that decompose complexity into predefined categories [71]. Because these frameworks are rule-based and widely standardized, they align closely with the strengths of generative AI in pattern recognition. The only exception is risk analysis, which belongs to the Strategy Evaluation (SE) domain. Overall, the evidence suggests the EA tasks exhibit a particularly strong alignment with generative AI tools and are likely to remain the most fertile ground for future development as perceived by the consultants.

In sum, the findings demonstrate that generative AI is already delivering tangible benefits in strategic planning, particularly in highly structured frameworks where consultants report strong satisfaction and consensus. Therefore, this cluster of activities can be labeled as mature regarding generative AI usage. At the same time, consultants recognize considerable promise in less standardized activities such as competitor, market, risk and trend analyses, though here current applications fall short of their expectations. This cluster is remarkable for its strong future potential. Moreover, consultants highlight that generative AI is increasingly used as a strategic sparring partner, supporting idea generation and client-focused brainstorming. Taken together, these results suggest that generative AI's benefits are threefold: it already enables support for structured analytical frameworks, it holds significant promise for more interpretive tasks and it provides immediate value as a creativity partner in client-oriented brainstorming. The next section turns to those strategic planning activities that are not yet ready to be effectively supported by generative AI: the challenges.

## 5.1.2 Challenges

### **Judgment-intensive activities as generative AI's current limitation**

The second part of the research question concerns the strategic planning activities where generative AI is not yet ready to provide meaningful support. A small set of activities perform considerably worse than others and they all share some common characteristics. Feasibility of a strategy, scenario creation and scenario evaluation are all activities that have both low satisfaction (average of activity means,  $M = 3.70$ ) and low importance ratings (average of activity means,  $M = 5.28$ ), combined with relatively high standard deviations (average of activity SDs,  $M = 1.44$ ). These results suggest that consultants are neither convinced of the present usefulness of generative AI in these areas nor optimistic about its future potential; in short, they are not yet ready to place trust in AI for these tasks. The high variation in satisfaction further indicates disagreement among consultants on whether generative AI is useful, reliable or even appropriate in such context. Unlike the "promising emergent activities", which reveal unmet expectations but high importance, these activities are characterized by a lack of confidence and limited perceived value. This pattern is further reflected in the gap scores. Although the mean gap (average of activity gaps,  $M = 1.58$ ) appears relatively high, this is largely a function of the very low satisfaction ratings: when baseline satisfaction is low, even small increases in perceived future importance produce a significant gap. More importantly, the relatively weak

importance ratings indicate that consultants have little faith in generative AI's ability to add meaningful value to these activities, either now or in the foreseeable future.

Taken together, these results identify feasibility assessments, scenario creation and scenario evaluation as the weakest areas of generative AI support. These activities share common characteristics that explain their poor performance with generative AI. Tasks such as assessing the feasibility of a strategy or creating and evaluating scenarios are judgment-intensive and require context-specific information, which is often not accessible to LLMs. Moreover, these tasks can be considered by consultants as mission-critical, making it difficult for them to delegate such crucial tasks to LLMs. They involve synthesizing ambiguous information, weighing competing alternatives and making decisions under uncertainty. Prompting for these activities also demands extensive preparation, as it is not as straightforward as a structured analysis like SWOT. For instance, evaluating the feasibility of a strategy requires detailed context-specific information together with criteria, which is both time-consuming and cognitively demanding. Consequently, consultants may perceive these activities as less valuable, since drafting appropriate prompts requires significantly more effort than in the case of structured frameworks. Unlike structured frameworks there are no standardized templates or universally accepted "right answers." Instead, these activities demand deep organizational knowledge, intuitive expertise and sensitivity to cultural factors in which LLMs do not yet excel in. In short, the very features that make these activities strategically significant also make them resistant to automation.

These activities thus require more expertise and the demographic specific scores (see See Figures E.1, E.2 and 4.20) reveal a clear divide. Senior and older consultants reported substantially higher gaps in these tasks (up to  $M = 2.4$ ), whereas junior and younger consultants reported much lower gaps (around  $M = 0.90$ ). This discrepancy is largely explained by satisfaction levels: senior consultants gave very low ratings (around  $M = 3.30$ ), with feasibility of a strategy even falling below  $M = 3.00$  ( $M = 2.54$ ). This pattern suggests that more experienced consultants, who have a deeper understanding of the complexity and stakes of these activities, are particularly skeptical of generative AI's value in this context. By contrast, younger consultants assigned similar importance scores but reported less skepticism, which may reflect a more limited appreciation of the strategic depth these tasks require. Notably, the Strategy & Operations department reported extremely low satisfaction alongside high importance, producing a significantly larger gap than other departments. Given that this department has the most expertise and direct experience with these activities, these results support the interpretation that less experienced consultants may overestimate generative AI's potential, as they do not yet fully recognize the expertise and contextual judgment that such tasks demand.

The open responses support this interpretation. As one consultants explained, "Gen AI will have a really hard time to really understand strategic planning on a practical level. (...) At the moment he can answer high level good, low level not good." Another warned that "sometimes GenAI uses hallucination and think of scenarios that are not possible." Others were concerned that such tools may create a false sense of reliability: "It can provide very convincing answers, however it makes you lazy to actually read what it is saying." These reflections reinforce the

quantitative findings: consultants are wary of delegating highly judgment-heavy and mission-critical tasks to a generative AI system that is prone to generic outputs, hallucinations and misplaced confidence in complex tasks. Yafei et al. highlight the severe risk of hallucinations of using LLMs. These risks are particularly severe in strategic planning, where hallucinated outputs may appear authoritative yet lack factual grounding, making them difficult to detect without expert oversight and potentially leading to misguided or incorrect decisions [86]. Thus, the danger is not only mediocre output, but the false authority of hallucinated results in high-stakes contexts.

Recent scholarship on generative AI and management also provides useful perspective. Raisch described the "automation-augmentation paradox," emphasizing that while generative AI systems excel at structured and repetitive tasks, complex managerial activities involving tacit knowledge and contextual judgment still require human judgment [67]. Feasibility assessments and scenario creation and evaluation fall directly into this category. These activities involve ambiguity and organizational dynamics, which cannot be reliably formalized for LLMs. In such cases, generative AI should be viewed as an augmentation of human expertise rather than a replacement. This distinction explains why consultants in the survey remain cautious: generative AI may provide supportive inputs, but it cannot assume responsibility for decision of such strategic weight.

Crucially, all of these activities belong to the domain of SE. As previously discussed SE is concerned with assessing whether a chosen or proposed strategy is feasible, consistent and aligned with organizational goals, often using criteria such as risk analysis, scenario evaluations and feasibility assessments [71]. By their nature, these tasks require interpreting uncertain signals weighing trade-offs and applying deep contextual knowledge processes that resist codification, so that they can be more easily used for generative AI. It is precisely this evaluative and judgment-heavy character that explains why generative AI performs poorly here, in stark contrast to its stronger fit with EA tools [67]. In short, while EA aligns with the structured, pattern-based strength of LLMs, SE remains beyond the current frontier of what generative AI can reliably support.

Taken together, the results indicate that feasibility assessments and scenario creation and evaluation represent strategic planning activities where generative AI is not yet suitable as a stand-alone tool. While younger consultants and the non-dedicated strategy planners show greater optimism, senior consultants and the daily practitioners, who carry responsibility for high-stakes outcomes, express stronger skepticism, which is reflected in their lower satisfaction ratings and higher gap scores. This demographic divide highlights that for judgment-intensive, mission-critical tasks, human expertise and interpretive capacity remain indispensable. At present, generative AI can play only a supplementary role in these activities, offering selective augmentation rather than full automation in these domains.

## 5.2 Prompting as a critical but underdeveloped competence

In addition to the main findings on benefits and challenges, the survey highlights a striking discrepancy between consultants' prompting practices and expressed needs. Figures 4.14 and 4.16 show an interesting contrast while an overwhelming majority of consultants agree that tailored prompting strategies are essential for effective use of generative AI in strategic planning (over 89% in agreement), only 4% currently employ prompts specifically designed for strategic planning activities. Instead, most either create a prompt from scratch (45.6%) or use whatever comes to mind in the moment (30.6%). This gap highlights that consultants recognize the strategic importance of prompting, but lack the structured skills or resources to implement it effectively and are therefore not yet mature in digital competence. As one respondent summarized: "Prompting is key, (...) then you will get most value out of Gen AI during strategic planning activities." In other words, awareness of prompting's importance is already widespread, but it has not yet been translated into systematic practice. This gap positions prompting as the bottleneck competence: consultants recognize its importance, but lack the skills to operationalize it.

The inconsistency of current practice is also visible in how consultants evaluate existing training. Figure 4.13 shows highly divergent views on the usefulness of recent AI training, with ratings distributed from "not at all useful" to "extremely useful." This variation suggests that while some sessions addressed consultants' needs, others remained too generic, introductory or insufficiently interactive, failing to deliver task-specific skills. When we compare this with the stronger consensus in favor of tailored prompting strategies, it becomes clear why consultants value generative AI training in principle but remain dissatisfied in practice. Much of the training they have received so far does not meet their specific needs. Consultants note that LLM may "hallucinate" and that it is "key to keep human in the loop", especially the human judgment. As one noted "another challenge is training and developing junior staff in also making this judgment on GenAI output," referring to remaining critical of the output of LLM. In short, prompting is already recognized as critical, but without targeted and tailored training, consultants remain unable to operationalize it effectively.

The demand for structured support is equally evident in Figure 4.15. When asked what would help them use generative AI more effectively, the most frequently selected option was "hands-on training in writing effective prompts (workshop)" (23.9%), followed closely by "prompting templates or examples" (19.6%). This demonstrates that consultants do not merely want access to generative AI tools, they want to learn how to master them through guided, interactive format. As one respondent put it, support should include "training or guidelines on how to evaluate AI output." This reflects that consultants seek not only to generate better prompts, but also to critically assess the quality and reliability of outputs. Put differently, consultants see value in generative AI only when paired with the competence to interrogate and refine its results through mastering prompting.

This finding resonates with the academic perspective on AI augmentation. Raisch and Krakowski emphasize that while generative AI can automate structured tasks, its value in

knowledge-intensive work depends heavily on how humans frame, contextualize and guide the system [67]. In this sense, prompting becomes more than a technical skill: it is the key mechanism through which consultants can enable generative AI to augment, rather than replace, their strategic expertise. The survey results suggest that without stronger prompting competence, consultants risk leaving much of generative AI's potential untapped.

In sum, the results suggest that prompting should be treated as a core digital competence in management consulting. Currently, the lack of tailored prompting strategies leads to unmet needs and untapped potential. Yet, there is broad consensus on the need for these strategies and strong demand for practical, workshop-based training. Developing this competence could accelerate generative AI adoption from ad-hoc experimentation toward systematic prompt engineering, establishing prompting as the bottleneck capability that determines whether generative AI delivers reliable strategic value. Prompting should therefore be understood as the critical lever that determines whether generative AI augments consultants effectively or remains underutilized.

### **5.3 Limitations**

Every research has its limitations and this study is no exception. The limitations are discussed below. This study followed an exploratory sequential mixed methods approach, combining initial qualitative input with a broader quantitative survey. While this design is appropriate for a novel and underdeveloped topic such as generative AI in strategic planning within management consulting, it also introduces several limitations.

The initial qualitative phase was constrained by selection bias, as the interview sample was limited to only six expert. While sufficient for theme generation in a homogeneous context, this small N restricts the diversity of perspectives, limiting the range of identified themes [36]. Furthermore, the semi-structured, researcher-led format poses a risk of confirmation bias and interpretive drift, despite the use of an interview guide. Since these interviews informed the subsequent survey, early framing effects may have unintentionally privileged certain strategic planning activities over others.

The survey phase introduced constraints related to sampling, response patterns and data characteristics. Self-selection bias arose from voluntary participation within a single Big Four advisory firm, likely attracting consultants already interested in generative AI. Therefore, the results may skew toward more favourable assessments. Though the survey was distributed across diverse roles and departments to capture a heterogeneous sample and mitigate the bias, the core limitation of external validity remains.

The data is inherently constrained by the self-report on the Likert scales, capturing subjective perception rather than objective performance. Despite anonymity, the risk of social desirability bias remains, with respondents potentially overstating positive attitudes due to its current hype or perceived professional expectations. This is particularly relevant given the unanimous optimism in the responses, which is therefore interpreted with caution. The potential for non-response

bias due to survey fatigue was mitigated through a pilot test and a consistent response scale, to reduce cognitive load.

Analytically, the exclusion of "Not applicable / Not familiar with" responses, while improving internal validity for those items, reduced the effective sample size and may have biased the findings toward the view of consultants most engaged in specific strategic planning activities.

The generalizability of the findings is further constrained by the sample's concentricity. Although various seniority levels, departments and function were reached, the overall sample size was modest and regionally concentrated within one firm. Moreover, the population was somewhat overrepresented by the technology department and senior roles. Therefore, the findings should be generalized with caution to the consulting industry at large.

The demonstration used a single model family (ChatGPT) and a single prompting workflow. Performance and outputs may vary across models, versions and across alternative prompting strategies. As such, the findings should not be overgeneralized to other tools or firms.

Taken together, although several mitigation were taken, these limitations constrain the external validity and generalizability of the findings, yet they do not undermine the study's contribution in providing early insights into the role of generative AI in strategic planning.

## **5.4 Future research**

Building on the limitations outlined above, several avenues for future research can be identified.

First, expanding the qualitative phase to include a larger and more diverse set of interviewees across different firms and sectors would strengthen the breadth and transferability of insights. Comparative studies between Big Four consultancies and boutique firms could highlight whether organizational size, culture or sectoral focus shapes perceptions of generative AI in strategic planning.

Second, the quantitative phase would benefit from a larger more diversified sample size. A multi-firm survey could reduce organizational bias and allow for benchmarking across industries. Moreover, complementing self-reported perception with objective performance metrics, such as efficiency gains, quality of deliverables or client outcomes, would enable more robust conclusions in terms of actual value.

Third, future studies should explore the role of different generative AI models and prompting strategies. The practical demonstration in this thesis was limited to ChatGPT; subsequent research could focus on one activity and systematically compare outputs across multiple models, version and configurations. To assess how tool choice and prompting technique affect outcome.

Fourth, future research could focus on the more challenging activities identified in this study. These tasks often require extensive preparation and carefully crafted prompts, making them resources-intensive. However, the emergence of live prompting techniques (i.e., LLMs listening to conversations) may open new opportunities for achieving effective outcomes with minimal effort. Investigating whether live prompting can reduce the cognitive and time burden

of designing prompts, while maintaining reliability and quality, would be a valuable direction for further exploration.

Finally, as prompting and generative AI are nascent concepts a longitudinal research is recommended. Tracking changes in consultant perceptions and practices over time would provide insight into adoption dynamics, shifts in trust and the long-term integration of generative AI strategic planning activities.

## Chapter 6 Conclusion

This research set out to explore where generative AI can and cannot support strategic planning activities in management consulting, in other words, its benefits and challenges. Using an exploratory sequential mixed methods research design, the study combined six qualitative expert interviews with consultants from KPMG and SAP, a quantitative survey with 128 responses from KPMG consultants and a practical demonstration to bridge the gap between theory and applied consulting practice.

Within the six-step strategic planning model, Environmental Analysis and Strategy Evaluation emerged as the most relevant phases for generative AI support. The subsequent survey conducted a gap analysis: respondents rated current satisfaction and future importance, with future importance rated significantly higher than current satisfaction, signaling potential that is not yet fully realized in practice.

The findings point to three main benefits. First, mature use cases: LLMs reliably assist structured analytical frameworks, such as SWOT, PESTLE and Porter's Five Forces, where the model's reasoning aligns with predefined logic and categories. Second, promising emergent activities, which are market, risk, competitor analyses and technological trend forecasting, show exceptionally high future importance despite their current immaturity, suggesting clear potential for future development. Third, the "sparring partner" role, in which consultants use generative AI to brainstorm over clients, exemplifies how the technology enhances creativity and accelerates idea generation without undermining human judgment.

The study also uncovers clear challenges that constrain broader adoption. Tasks involving feasibility assessment, scenario creation and scenario evaluation remain difficult for current models due to their context-dependent and judgment-intensive nature. These activities require well-crafted prompts. The quantitative findings suggest that prompting literacy, the ability to design clear, structured and context-aware prompts, has become a critical capability for consultants. Consultants, however, note that this emerging critical digital competence is underdeveloped.

The practical demonstration complements these insights by showing how deliberate prompt engineering can systematically improve output quality. Through six prompting stages, ranging from zero-shot baseline to self-refining reasoning overlays, the demonstration illustrates the output of LLMs can evolve.

This thesis demonstrates that generative AI already supports various strategic planning activities, while limitations in contextual understanding and judgment constrain its applicability in others. By clarifying where generative AI provides tangible value and where it does not, this study establishes a foundation for future research and informs areas of practical exploration for consultants.

# Bibliography

- [1] Josh Achiam et al. “Gpt-4 technical report”. In: *arXiv preprint arXiv:2303.08774* (2023).
- [2] F.J. Aguilar. *Scanning the business environment*. Macmillan, 1967. URL: <https://books.google.nl/books?id=sn1EAAAAIAAJ>.
- [3] Alaa Althubaiti. “Information bias in health research: definition, pitfalls, and adjustment methods”. In: *Journal of multidisciplinary healthcare* (2016), pp. 211–217.
- [4] H Igor Ansoff et al. “Strategies for diversification”. In: *Harvard business review* 35.5 (1957), pp. 113–124.
- [5] Helen L. Ball. “Conducting Online Surveys”. In: *Journal of Human Lactation* 35.3 (2019). PMID: 31084575, pp. 413–417. DOI: 10.1177/0890334419848734. eprint: <https://doi.org/10.1177/0890334419848734>. URL: <https://doi.org/10.1177/0890334419848734>.
- [6] Housseem Eddine Ben Messaoud. “A Review on the Importance of Strategic Planning in Business”. In: (July 2022), pp. 1–5. DOI: 10.14445/23939125/IJEMS-V9I7P101.
- [7] Mostafa Ali Benzaghta et al. “SWOT analysis applications: An integrative literature review”. In: *Journal of Global Business Insights* 6.1 (2021), pp. 54–72.
- [8] Carolyn Boyce and Palena Neale. *Conducting in-depth interviews: A guide for designing and conducting in-depth interviews for evaluation input*. Vol. 2. Pathfinder international Watertown, MA, 2006.
- [9] Peter Brews and Devavrat Purohit. “Strategic Planning in Unstable Environments”. In: *Long Range Planning* 40.1 (2007), pp. 64–83. ISSN: 0024-6301. DOI: <https://doi.org/10.1016/j.lrp.2006.12.001>. URL: <https://www.sciencedirect.com/science/article/pii/S0024630106001221>.
- [10] Tom Brown et al. “Language models are few-shot learners”. In: *Advances in neural information processing systems* 33 (2020), pp. 1877–1901.
- [11] Alan Bryman. “Integrating quantitative and qualitative research: how is it done?” In: *Qualitative research* 6.1 (2006), pp. 97–113.
- [12] John Bryson and Bert George. *Strategic Planning for Public and Nonprofit Organizations: A Guide to Strengthening and Sustaining Organizational Achievement, 6th Edition*. July 2024. ISBN: 978-1-394-27402-4.
- [13] John M Bryson. “Strategic planning for public and nonprofit organizations”. In: *International Encyclopedia of the Social & Behavioral Sciences: Second Edition*. Elsevier Inc., 2015, pp. 515–521.

- [14] Kiran Busch et al. “Just tell me: Prompt engineering in business process management”. In: *International Conference on Business Process Modeling, Development and Support*. Springer. 2023, pp. 3–11.
- [15] Anna Canato and Antonio Giangreco. “Gurus or Wizards? A Review of the Role of Management Consultants”. In: *European Management Review* 8 (Dec. 2011). DOI: 10.1111/j.1740-4762.2011.01021.x.
- [16] Banghao Chen et al. “Unleashing the potential of prompt engineering for large language models”. In: *Patterns* (2025).
- [17] John W Creswell and Vicki L Plano Clark. *Designing and conducting mixed methods research*. Sage publications, 2017.
- [18] John W Creswell and J David Creswell. *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications, 2017.
- [19] Paul G. Curran. “Methods for the detection of carelessly invalid responses in survey data”. In: *Journal of Experimental Social Psychology* 66 (2016). Rigorous and Replicable Methods in Social Psychology, pp. 4–19. ISSN: 0022-1031. DOI: <https://doi.org/10.1016/j.jesp.2015.07.006>. URL: <https://www.sciencedirect.com/science/article/pii/S0022103115000931>.
- [20] Christina Davidson. “Transcription: Imperatives for qualitative research”. In: *International journal of qualitative methods* 8.2 (2009), pp. 35–52.
- [21] Saraswati Dawadi, Sagun Shrestha, and Ram A Giri. “Mixed-methods research: A discussion on its types, challenges, and criticisms”. In: *Journal of Practical Studies in Education* 2.2 (2021), pp. 25–36.
- [22] Robert F DeVellis and Carolyn T Thorpe. *Scale development: Theory and applications*. Sage publications, 2021.
- [23] Jacob Devlin et al. “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”. In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Ed. by Jill Burstein, Christy Doran, and Tamar Solorio. Minneapolis, Minnesota: Association for Computational Linguistics, June 2019, pp. 4171–4186. DOI: 10.18653/v1/N19-1423. URL: <https://aclanthology.org/N19-1423/>.
- [24] Dmytro Dosyn et al. *Application of the analytical hierarchy process to evaluate hype AI technologies*. 2024.
- [25] Christopher P Durney and Richard G Donnelly. “Managing the effects of rapid technological change on complex information technology projects”. In: *Journal of the Knowledge Economy* 6 (2015), pp. 641–664.
- [26] T Dybå and Nils Brede Moe. “Rethinking the concept of software process assessment”. In: *European Software Process Improvement Conference (EuroSPI), Pori, Finland*. 1999.

- [27] Peter Emerson. “The original Borda count and partial voting”. In: *Social Choice and Welfare* 40.2 (2013), pp. 353–358.
- [28] Angela Fan, Mike Lewis, and Yann Dauphin. “Hierarchical Neural Story Generation”. In: Jan. 2018, pp. 889–898. DOI: 10.18653/v1/P18-1082.
- [29] Denis Federiakin et al. “Prompt engineering as a new 21st century skill”. In: *Frontiers in Education*. Vol. 9. Frontiers Media SA. 2024, p. 1366434.
- [30] Michael D. Fetters, Leslie A. Curry, and John W. Creswell. “Achieving Integration in Mixed Methods Designs—Principles and Practices”. In: *Health Services Research* 48.6pt2 (2013), pp. 2134–2156. DOI: <https://doi.org/10.1111/1475-6773.12117>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/1475-6773.12117>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/1475-6773.12117>.
- [31] Kraig Finstad. “Response interpolation and scale sensitivity: Evidence against 5-point scales”. In: *Journal of usability studies* 5.3 (2010), pp. 104–110.
- [32] R Edward Freeman. *Strategic management: A stakeholder approach*. Cambridge university press, 2010.
- [33] Fiona Fylan. “Semi-structured interviewing”. In: *A handbook of research methods for clinical and health psychology* 5.2 (2005), pp. 65–78.
- [34] Daniel Mandel Gandrita. “Improving strategic planning: The crucial role of enhancing relationships between management levels”. In: *Administrative Sciences* 13.10 (2023), p. 211.
- [35] Bert George, Richard M. Walker, and Joost Monster. “Does Strategic Planning Improve Organizational Performance? A Meta-Analysis”. In: *Public Administration Review* 79.6 (2019), pp. 810–819. DOI: <https://doi.org/10.1111/puar.13104>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/puar.13104>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/puar.13104>.
- [36] Greg Guest, Arwen Bunce, and Laura Johnson. “How many interviews are enough? An experiment with data saturation and variability”. In: *Field methods* 18.1 (2006), pp. 59–82.
- [37] Daniel S. Hain et al. “A text-embedding-based approach to measuring patent-to-patent technological similarity”. In: *Technological Forecasting and Social Change* 177 (2022), p. 121559. ISSN: 0040-1625. DOI: 10.1016/j.techfore.2022.121559. URL: <https://www.sciencedirect.com/science/article/pii/S0040162522000919>.
- [38] Robert L. Harrison. “Using mixed methods designs in the Journal of Business Research, 1990–2010”. In: *Journal of Business Research* 66.11 (2013), pp. 2153–2162. ISSN: 0148-2963. DOI: <https://doi.org/10.1016/j.jbusres.2012.01.006>. URL: <https://www.sciencedirect.com/science/article/pii/S0148296312000252>.

- [39] Monique M Hennink, Bonnie N Kaiser, and Vincent C Marconi. “Code saturation versus meaning saturation: how many interviews are enough?” In: *Qualitative health research* 27.4 (2017), pp. 591–608.
- [40] Nataliya V. Ivankova, John W. Creswell, and Sheldon L. Stick. “Using Mixed-Methods Sequential Explanatory Design: From Theory to Practice”. In: *Field Methods* 18.1 (2006), pp. 3–20. DOI: 10.1177/1525822X05282260. eprint: <https://doi.org/10.1177/1525822X05282260>. URL: <https://doi.org/10.1177/1525822X05282260>.
- [41] Dawid Jarco and Lukasz Sulkowski. “Is ChatGPT Better at Business Consulting than an Experienced Human Analyst? An Experimental Comparison of Solutions to a Strategic Business Problem”. In: *Forum Scientiae Oeconomia* 11.2 (June 2023), pp. 87–109. DOI: 10.23762/FSO\_VOL11\_N02\_4. URL: <https://ojs.wsb.edu.pl/index.php/fso/article/view/952>.
- [42] R Burke Johnson. “Examining the validity structure of qualitative research.” In: *Education* 118.2 (1997).
- [43] Feroze Kaliyadan and Vinay Kulkarni. “Types of variables, descriptive statistics, and sample size”. In: *Indian dermatology online journal* 10.1 (2019), pp. 82–86.
- [44] Ryan Idris Kamaruddin. “ChatGPT and the Future of Management Consulting: Opportunities and Challenges Ahead”. Doctoral dissertation. Massachusetts Institute of Technology, 2023. URL: <https://hdl.handle.net/1721.1/151586>.
- [45] Mikko Ketokivi and Xavier Castaner. “Strategic planning as an integrative device”. In: *Administrative Science Quarterly* 49.3 (2004), pp. 337–365.
- [46] Pawel Korzynski et al. “Artificial intelligence prompt engineering as a new digital competence: Analysis of generative AI technologies such as ChatGPT”. In: *Entrepreneurial Business and Economics Review* 11.3 (2023), pp. 25–37.
- [47] Milan Kubr. “Management consulting: A guide to the profession”. In: *International Labour Organization* (2002).
- [48] Chieh-Jung Liu. “Strategic environmental analysis-selected tools”. In: (2015).
- [49] Pengfei Liu et al. “Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing”. In: *ACM computing surveys* 55.9 (2023), pp. 1–35.
- [50] Dieuwertje Luitse and Wiebke Denkena. “The great Transformer: Examining the role of large language models in the political economy of AI”. In: *Big Data & Society* 8.2 (2021), p. 20539517211047734. DOI: 10.1177/20539517211047734. URL: <https://doi.org/10.1177/20539517211047734>.
- [51] James R Maina. “A Survey of Strategic Planning Practices of Management Consultancy Firms in Kenya”. Doctoral dissertation. University of Nairobi, 2007.

- [52] Aubrey L Mendelow. “Environmental scanning—the impact of the stakeholder concept”. In: *Second International Conference on Information Systems* (1981).
- [53] AI Meta. “Introducing meta llama 3: The most capable openly available llm to date”. In: *Meta AI* (2024).
- [54] H. Mintzberg et al. *Strategy Safari: A Guided Tour Through The Wilds of Strategic Mangament*. Free Press, 2005. ISBN: 9780743270571. URL: <https://books.google.nl/books?id=z0MIuP4ZS5gC>.
- [55] Henry Mintzberg et al. “The fall and rise of strategic planning”. In: *Harvard business review* 72.1 (1994), pp. 107–114.
- [56] Sarah K Morsbach and Ronald J Prinz. “Understanding and improving the validity of self-report of parenting”. In: *Clinical child and family psychology review* 9 (2006), pp. 1–21.
- [57] Muhammad Iqbal Muhammad Fayaz Muhammad Amin. “Artificial Intelligence in Decision-Making: A Descriptive Analysis of Its Impact on Strategic Planning”. In: *Journal of Business and Management Research* 3.2 (Sept. 2024), pp. 222–235. URL: <https://jbmr.com.pk/index.php/Journal/article/view/228>.
- [58] Helen Noble and Joanna Smith. “Issues of validity and reliability in qualitative research”. In: *Evidence-based nursing* 18.2 (2015), pp. 34–35.
- [59] Paul Nutt. “Making Strategic Choices”. In: *Journal of Management Studies* 39 (Dec. 2002), pp. 67–96. DOI: 10.1111/1467-6486.00283.
- [60] OpenAI. *ChatGPT (October 2025 version)*. Online. San Francisco, CA: OpenAI. 2025. URL: <https://chat.openai.com/> (visited on 10/25/2025).
- [61] Daniel M Oppenheimer, Tom Meyvis, and Nicolas Davidenko. “Instructional manipulation checks: Detecting satisficing to increase statistical power”. In: *Journal of experimental social psychology* 45.4 (2009), pp. 867–872.
- [62] Jean-Francois Orsini. “Artificial intelligence: A way through the strategic planning crisis?”. In: *Long Range Planning* 19.4 (1986), pp. 71–77. ISSN: 0024-6301. DOI: [https://doi.org/10.1016/0024-6301\(86\)90273-6](https://doi.org/10.1016/0024-6301(86)90273-6). URL: <https://www.sciencedirect.com/science/article/pii/0024630186902736>.
- [63] Dimple Patil, Nitin Rane, and Jayesh Rane. “Emerging and future opportunities with ChatGPT and generative artificial intelligence in various business sectors”. In: Oct. 2024, pp. 242–293. ISBN: 978-81-981367-8-7. DOI: 10.70593/978-81-981367-8-7\_6.
- [64] Michael E Porter. *How competitive forces shape strategy*. Springer, 1989.
- [65] Alec Radford et al. “Language models are unsupervised multitask learners”. In: *OpenAI blog* 1.8 (2019), p. 9.

- [66] Mohaimenul Azam Khan Raiaan et al. “A Review on Large Language Models: Architectures, Applications, Taxonomies, Open Issues and Challenges”. In: *IEEE Access* 12 (2024), pp. 26839–26874. DOI: 10.1109/ACCESS.2024.3365742.
- [67] Sebastian Raisch and Sebastian Krakowski. “Artificial Intelligence and Management: The Automation–Augmentation Paradox”. In: *Academy of Management Review* 46 (Jan. 2021), pp. 192–210. DOI: 10.5465/amr.2018.0072.
- [68] Laria Reynolds and Kyle McDonell. “Prompt programming for large language models: Beyond the few-shot paradigm”. In: *Extended abstracts of the 2021 CHI conference on human factors in computing systems*. 2021, pp. 1–7.
- [69] Aleksi Roima. “The disruption of the IT consulting business by large language models: an analysis of implications”. In: (2023).
- [70] Pranab Sahoo et al. *A Systematic Survey of Prompt Engineering in Large Language Models: Techniques and Applications*. Feb. 2024. DOI: 10.13140/RG.2.2.13032.65286.
- [71] D. Schendel and C.W. Hofer. *Strategic Management: A New View of Business Policy and Planning*. Little, Brown, 1979. ISBN: 9780316773126. URL: <https://books.google.nl/books?id=TRMPAQAAMAAJ>.
- [72] Sander Schulhoff et al. “Ignore This Title and HackAPrompt: Exposing Systemic Vulnerabilities of LLMs Through a Global Prompt Hacking Competition”. In: *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*. Ed. by Houda Bouamor, Juan Pino, and Kalika Bali. Singapore: Association for Computational Linguistics, Dec. 2023, pp. 4945–4977. DOI: 10.18653/v1/2023.emnlp-main.302. URL: <https://aclanthology.org/2023.emnlp-main.302/>.
- [73] Sander Schulhoff et al. “The prompt report: A systematic survey of prompting techniques”. In: 5 (June 2024).
- [74] Veis Šerifi and Predrag Dašić. “Characteristics of traditional and contemporary models of organizational structures”. In: *7th International Conference ICQME*. 2012, pp. 333–339.
- [75] Christoph Johann Stettina and Werner Heijstek. “Necessary and neglected? An empirical study of internal documentation in agile software development teams”. In: *Proceedings of the 29th ACM international conference on Design of communication*. 2011, pp. 159–166.
- [76] Valerie M Sue and Lois A Ritter. *Conducting online surveys*. Sage, 2007.
- [77] Mohsen Tavakol and Reg Dennick. “Making sense of Cronbachs alpha”. In: *International journal of medical education* 2 (2011), p. 53.
- [78] LearnLM Team et al. *LearnLM: Improving Gemini for Learning*. 2024. arXiv: 2412.16429 [cs.CY]. URL: <https://arxiv.org/abs/2412.16429>.
- [79] Martine Van Selm and Nicholas W Jankowski. “Conducting online surveys”. In: *Quality and quantity* 40 (2006), pp. 435–456.

- [80] Ashish Vaswani et al. “Attention Is All You Need”. In: *CoRR* abs/1706.03762 (2017). arXiv: 1706.03762. URL: <http://arxiv.org/abs/1706.03762>.
- [81] Jason Wei et al. “Chain-of-thought prompting elicits reasoning in large language models”. In: *Advances in neural information processing systems* 35 (2022), pp. 24824–24837.
- [82] Richard Whittington. “Completing the practice turn in strategy research”. In: *Organization studies* 27.5 (2006), pp. 613–634.
- [83] Carola Wolf and Steven W Floyd. “Strategic planning research: Toward a theory-driven agenda”. In: *Journal of management* 43.6 (2017), pp. 1754–1788.
- [84] Bill Wooldridge, Torsten Schmid, and Steven W Floyd. “The middle management perspective on strategy process: Contributions, synthesis, and future research”. In: *Journal of management* 34.6 (2008), pp. 1190–1221.
- [85] Tianyu Wu et al. “A Brief Overview of ChatGPT: The History, Status Quo and Potential Future Development”. In: *IEEE/CAA Journal of Automatica Sinica* 10 (2023), pp. 1122–1136. DOI: 10.1109/JAS.2023.123618. URL: <https://doi.org/10.1109/JAS.2023.123618>.
- [86] Xiang Yafei et al. “Generative AI in Industrial Revolution: A Comprehensive Research on Transformations, Challenges, and Future Directions”. In: *Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online)* 3.2 (2024), pp. 11–20. DOI: 10.60087/jklst.vol.3n2.p20. URL: <https://doi.org/10.60087/jklst.vol.3n2.p20>.

# Chapter A Interview guide

## Introduction

- First, I would like to ask if I may record this interview and use the data for my thesis. The data will be anonymized by name but not by function.
- Thank you for being open to this interview for my thesis.
- In my thesis, I want to investigate the pros and cons of using Generative AI for strategic planning activities within the management consulting sector. I am particularly interested in the specific activities carried out in daily practice, rather than a high-level overview and what effect Generative AI can have on these activities. I will later provide definitions of Generative AI and strategic planning.
- I define management consulting as the sector that advises companies on strategic, operational and technological issues.
- I will start by asking a few general questions.

## General Questions

- **Question:** Could you briefly tell me about yourself, including your role, activities and responsibilities?
- **Question:** Could you describe a typical workday? Please be as detailed as possible.

## Strategic Planning (Without Context)

- **Question:** Could you describe what you think strategic planning is?
  - *If no:* I understand that it's difficult to describe directly. Could you tell me how your organization (or department) reaches a decision about long-term goals and how you contribute to this?
  - How do you think organizations make decisions about what they want to achieve and how they can achieve it?
- **Question:** What are the daily activities or tasks that, consciously or unconsciously, directly or indirectly contribute to strategic planning? Think about specific activities at both an operational and strategic level.

## Definition of Strategic Planning

- To properly answer the next questions, I need to first explain a few terms.
- Strategic planning is a structured process consisting of 6 steps: setting goals, performing an environmental analysis, formulating a strategy, evaluating it, implementing it in the organization and finally, iteratively controlling and monitoring.

## The 6 Steps of Strategic Planning

1. **Setting Goals:** During this step, long-term goals (approx. 1.5 to 3 years into the future) are defined. These goals are often numerical and specific, such as increasing market share by 15% or annual revenue by 20%.
2. **Environmental Analysis:** In this phase, the organization performs various analyses, such as a SWOT analysis, to identify the strengths, weaknesses, opportunities and threats. This helps an organization determine where opportunities lie.
3. **Strategy Formulation:** Based on the information gathered in the previous step, the organization decides which direction it wants to go and how to get there. This step has 5 sub-steps: identifying the problem, analyzing the problem, developing various strategies to tackle the problem and achieve the goals, evaluating the alternatives for suitability and feasibility and choosing the most suitable strategy. This formal, structured process ensures that issues can be resolved.
4. **Strategy Evaluation:** This step is essential for determining whether the strategy is achieving the right goals, given the environment the organization is in. This can be tested using tools like risk analysis, early performance indicators and return on investment.
5. **Strategy Implementation:** This step divides the strategy into “actionable steps,” which are subtasks that can be easily implemented in the company. It is important to adapt the organizational structure, processes and behavior across the entire organization to this strategy. This involves clear roles and responsibilities and the proper allocation of resources.
6. **Strategy Control and Monitoring:** During this phase, progress is checked to ensure that goals are being achieved. Key Performance Indicators (KPIs) are established, which can be derived from the early performance indicators. It remains important to continuously evaluate the strategy so that adjustments can be made if necessary.

## Activities of Strategic Planning

- **Question:** Now that you have heard this explanation, what activities in your daily work have an interface with the definition of strategic planning just mentioned?

- *(Think about data analysis, reporting, customer interaction, problem identification.)*
- *If no:* I understand that you may not immediately recognize these activities in your role. Could you describe how your organization handles these tasks or decisions? Who normally does that and how do you get the information or direction for your own work to contribute?
- **Question:** In what ways do you contribute to your organization’s long-term goals or strategy? Can you provide an example of how you do this in your work?
- **Question:** What specific activities are performed to operationalize and implement a strategy in your organization?
- **Question:** How do you monitor the progress of projects or initiatives you are working on? When is something successful?
- **Question:** Is AI already involved in this or would that be useful?
- **Question:** Could you describe how data analysis and collection are part of your work and how you think they contribute to strategic planning?
- **Question:** Can you describe how reports are created in your organization and what role they play in strategic decision-making?

## Generative AI

- Generative AI is an emerging technology that can create new content based on existing data. This can include music, visual material (photos and videos) and, most famously, text. In this study, I am specifically focusing on the textual side of Generative AI: “**conversational Large Language Models**” (LLMs).
- These are advanced computer models, trained on massive amounts of text, that not only respond like a search engine but are also capable of making logical connections, holding conversations and remembering the context of previous input. They can also search internal files, making it easier to retrieve and combine information.
- Popular examples of conversational LLMs are ChatGPT (developed by OpenAI) and Gemini (from Google). They generate text by predicting which word or phrase logically follows the given prompt or earlier messages in the conversation. This allows them to formulate coherent and contextually appropriate answers that sometimes sound surprisingly human.
- At the same time, there are also risks in using these models. For example, they can provide inaccurate or outdated information (so-called “**hallucinations**”) or adopt certain biases from their training data. For strategic planning activities, it is therefore important to consider both the possible benefits and the limitations.

- **Question:** For what purposes are you currently using Generative AI tools in your daily work?
  - *If no:* what is the reason you are not using Generative AI tools?
- **Question:** Looking at the six steps of strategic planning
  - Goal Setting
  - Environmental Analysis
  - Strategy Formulation
  - Strategy Evaluation
  - Strategy Implementation
  - Strategy Control & Monitoring

could you give a top three of the steps where Generative AI could have a positive impact on the respective activities and explain for each step why you chose it?

- **Question:** Do you have suggestions or ideas about how and where Generative AI could be further integrated into daily strategic planning activities?
- **Question:** Do you think there are specific human skills or competencies that Generative AI (still) cannot take over, but that are important in strategic planning? If so, which ones? (For example, stakeholder management, intuitive decision-making.)
- **Question:** Are there any strategic planning activities where Generative AI would not be suitable, apart from ethical or legal considerations?

## Conclusion

- **Question:** Do you have any other insights you would like to share about strategic planning and AI in your work?

# Chapter B Survey instrument

*Hi colleague, thank you for contributing to this research. The survey will take between 10-15 minutes. Your input is greatly appreciated!*

*This research explores how Generative AI supports strategic planning activities in management consulting. Your input helps shape KPMG's wider AI journey and contributes to a tailored prompting strategy to help you get more out of tools like ChatGPT in your daily work.*

## **Definitions: Generative AI & Strategic Planning Practices**

- In this study, Generative AI refers to tools like ChatGPT, Copilot or internal KPMG AdvisoryGPT that use artificial intelligence to create new content, such as text, images or summaries. These tools are trained on large datasets and can support tasks like drafting analyses, brainstorming ideas and summarizing reports.
- Strategic Planning Practices are structured, iterative activities that help organizations set long-term goals, analyze their environment, choose strategies, allocate resources and track progress. These activities occur in different phases within projects. This research focuses on two key steps:
  - **Environmental Analysis:** gathering and interpreting data on markets, competitors, stakeholders and trends to identify opportunities.
  - **Strategy Evaluation:** assessing whether a strategy is realistic, goal-aligned and risk-aware.

## **What I ask of you**

- Please read each question and rate on a 7-point Likert scale. Honest answers help generate meaningful insights.

## **Privacy**

- All responses are anonymous, used only for research and stored securely.

## **Declaration of consent**

- I have read the information provided, understand the definitions of Generative AI and strategic planning and agree to participate in this study.

- I confirm that I am currently working in a consulting-related role, have read the informed consent and agree to participate in this research.

## **Introduction Questions**

The following demographic items are used to assess seniority and career profession. Please respond based on your current role, experience level and age group.

### **What is your job title?**

- Consultant
- Senior consultant
- Manager
- Senior manager
- Director
- Partner

### **How many years of experience do you have in consulting?**

- 0-2 years
- 3-7 years
- 8-15 years
- 15+ years

### **What department are you part of?**

- Technology
- Risk & regulatory
- Finance & business services
- Deals
- Strategy & operations

### **What is your age (in years)?**

- 18-21
- 22-34
- 35-49
- 50-65
- 65+

**How often do you actively use Generative AI in your professional work?**

- Never
- Rarely
- Monthly
- Weekly
- Daily

**How often do you perform strategic-planning work?**

- Never
- Rarely
- Monthly
- Weekly
- Daily

**Usage of Generative AI**

*Remember, there are no right or wrong answers; I am interested in your honest opinion. The statements provided are hypothetical. Please consider each statement thoughtfully and provide your insights.*

- **I am satisfied with the current usefulness of the Generative AI in my professional work.**
- **I think Generative AI will be important in the future in my professional work.**
- **I find Generative AI easy to use in my professional work.**
- **I feel that using Generative AI at work could pose potential risks.**

**Strategic Planning Practices: Environmental Analysis**

*The following questions relate to the strategic planning practices of environmental analysis. For each question, the specific practice (e.g. SWOT analysis) is shown in the top-left corner. There will be two questions for each practice, one testing the current satisfaction and the other the future importance of Generative AI in that practice. Please try to answer each question as accurately as possible based on your experience and expectations.*

- **SWOT Analysis** - Identify strengths, weaknesses, opportunities, threats to inform strategy.
  - I am satisfied with the current usefulness of Generative AI in supporting SWOT analysis.

- I believe Generative AI will be important in the future for SWOT analysis.
- **PESTLE analysis** - Scan political, economic, social, technological, legal, environmental factors.
  - I am satisfied with the current usefulness of Generative AI in supporting PESTLE analysis.
  - I believe Generative AI will be important in the future for PESTLE analysis.
- **Porter's Five Forces** - Assess industry rivalry, substitutes, buyers, suppliers, entry threats.
  - I am satisfied with the current usefulness of Generative AI in supporting Porter's Five Forces.
  - I believe Generative AI will be important in the future for Porter's Five Forces.
- **Market analysis** - Collect and analyse market size, segments, growth and competitors.
  - I am satisfied with the current usefulness of Generative AI in market analysis.
  - I believe Generative AI will be important in the future for market analysis.
- **Competitor analysis** - Evaluating rivals' strengths, weaknesses, strategies and performance.
  - I am satisfied with the current usefulness of Generative AI in competitor analysis.
  - I believe Generative AI will be important in the future for competitor analysis.
- **Technological market trends** – Project future trends and technologies affecting markets and strategy.
  - I am satisfied with the current usefulness of Generative AI in trend and technology forecasting.
  - I believe Generative AI will be important in the future for trend and technology forecasting.
- **Brainstorming on client needs** – Generate ideas addressing client needs, issues, aspirations and opportunities.
  - I am satisfied with the current usefulness of Generative AI in brainstorming on client needs.
  - I believe Generative AI will be important in the future for brainstorming on client needs.
- **Stakeholder analysis** – Identify, map, prioritise stakeholders' interests, influence and expectations.

- I am satisfied with the current usefulness of Generative AI in stakeholder analysis.
- I believe Generative AI will be important in the future for stakeholder analysis.
- **AS-IS analysis of current organizational situation** – Describe current processes, structure and performance to establish baseline for improvement.
  - I am satisfied with the current usefulness of Generative AI for making an AS-IS analysis of the current organizational situation.
  - I believe Generative AI will be important in the future for making an AS-IS analysis of the current organizational situation.

### **Strategic Planning Practices: Strategy Evaluation**

*The following questions relate to the strategic planning step of strategy evaluation.*

- **Strategic scenario creation** – Imagine different possible future situations for the business.
  - I am satisfied with the current usefulness of Generative AI in scenario creation.
  - I believe Generative AI will be important in the future for scenario creation.
- **Strategic scenario evaluation** – Systematically analyzing scenarios to assess strategic viability and implications.
  - I am satisfied with the current usefulness of Generative AI in scenario evaluation.
  - I believe Generative AI will be important in the future for scenario evaluation.
- **Peer-review synthesis** – Colleagues, external parties or tools review strategic plan and suggest clear improvements.
  - I am satisfied with the current usefulness of Generative AI in peer review synthesis.
  - I believe Generative AI will be important in the future for peer review synthesis.
- **Risk analysis** – Identify potential risks, analyse causes, consequences and mitigations.
  - I am satisfied with the current usefulness of Generative AI in risk analysis.
  - I believe Generative AI will be important in the future for risk analysis.
- **Risk probability assessment** – Estimate probability and impact scores for each identified risk.
  - I am satisfied with the current usefulness of Generative AI in risk probability assessment.
  - I believe Generative AI will be important in the future for risk probability assessment.
- **Brainstorming on strategy feasibility** – Can we realistically execute this strategy?.

- I am satisfied with the current usefulness of Generative AI in brainstorming on the feasibility of a strategy.
- I believe Generative AI will be important in the future for brainstorming on the feasibility of a strategy.

## **Personal Experiences and Reflections**

*In this final section, we ask for your personal experiences and reflections. Your responses will help us understand how Generative AI is used in practice during strategic planning. Feel free to mention both positive and negative aspects and provide concrete examples where possible.*

- **What do you think about integrating Generative AI in strategic planning activities?**
- **Can you describe a situation, if any, where Generative AI supported you in a strategic planning activity as mentioned in the survey? What worked well?**
- **Describe one limitation or risk you experienced when using Generative AI for strategic-planning activities.**
- **Have you received any internal or external training on the use of Generative AI?**
  - Yes, internal training at KPMG
  - Yes, external training (e.g. online course, university, external company, etc.)
  - No, I haven't received any training
- **What type of training have you received on using Generative AI for professional work?**
- **How useful was the most recent AI training you received?**
- **I would benefit from tailored prompting strategies for specific strategic planning activities.**
- **What kind of training, tools or support would help you use Generative AI more effectively in strategic planning?**
- **Which of the following prompting techniques are you familiar with? (Check all that apply)**
  - Ask Me Anything (AMA): Combines prompts and answers for richer model output.
  - Zero-Shot Prompting: No examples, model relies on pre-trained knowledge.
  - Few-Shot Prompting: Uses examples to guide model's generated responses.
  - Chain-of-Thought (CoT): Encourages step-by-step reasoning for complex tasks.

- Automatic Chain-of-Thought (Auto-CoT): Automates reasoning steps using structured prompt instructions.
  - Retrieval-Augmented Generation (RAG): Retrieves external data before generating final response.
  - None of the above
- **Which of the following best describes how you use prompts when working with Generative AI tools?**
    - I create my own prompts from scratch.
    - I reuse prompts I found online or from others.
    - I have prompts tailored to specific strategic tasks (e.g., SWOT, risk analysis).
    - I use whatever comes to mind in the moment.
    - I don't use any specific prompts.
    - Other:
  - **If you have found any specific prompts to be particularly effective when using Generative AI for strategic planning tasks, please feel free to share an example below (e.g., a prompt you used for competitor analysis or generating strategic alternatives).**

## B.1 Open-ended question answers

Table B.1: Respondents' perspectives on integrating Generative AI in strategic planning activities.

---

### What do you think about integrating Generative AI in strategic planning activities?

---

Might be good to collect all necessary information and scenarios. When it comes to making choices, the human factor becomes more important.

I think it will be useful, but I also feel a certain cautiousness about using it for strategic planning—the context is so important in many cases and hard to define perfectly in a prompt (at least for me right now).

Prompting is key. When you are able to create a well-designed prompt, you will get most value out of GenAI during strategic planning activities.

It will take over all the planning activities.

I believe it can support in getting more complete ideas, thinking out of the box, etc. The evaluation is, I guess, less applicable.

---

**Table B.1 – continued from previous page**

**What do you think about integrating Generative AI in strategic planning activities?  
(continued)**

As GenAI mainly looks at the current state of things and strategic planning is about the future state, I think GenAI can mainly help in drafting suggestions, but the work remains human.

It could be helpful, but people should not be depending on it.

It could support activities. Key in my view for strategic planning is: (a) to collect and structure a lot of data, which GAI could be very helpful with; (b) to assess, value, interpret the data and insights and decide on new, strategic plans. For this you need out-of-the-box thinking and new relations, which are not there yet. Therefore, I expect GAI to be less useful.

It should be integrated carefully, considering its biases.

It has immense opportunities to support strategic planning—identify and analyze a lot of information—but the strategic direction will and should still be owned and shaped by people. Strategizing is synthesizing large amounts of input to establish a direction. AI is vital nowadays for this.

It can definitely help while brainstorming, as it generates ideas that can be the starting point for new directions.

I believe generative AI can be of great help in terms of speed. With good input, you can get output that does not need much change anymore. This means that more work can get done.

We overestimate the impact.

It has the potential to enhance decision-making by providing new perspectives, generating creative ideas and automating certain tasks. These AI tools can analyze vast amounts of data quickly, identify patterns and help in scenario planning.

Useful to generate ideas, but since GenAI draws from past text and extrapolates, it also hallucinates. So far GenAI also has a time limit on past data in its source. It is key to keep the human in the loop to avoid logical fallacies and ensure conclusions remain relevant.

It will be challenging to integrate GenAI in client situations, as it generally lacks the internal context and dynamics of an organization.

I think this is extremely useful and can save a lot of time. It is, however, key that experienced professionals still review and make sure the output is correct and useful. Another challenge is training and developing junior staff to make this judgment as well.

Wise to use as an additional perspective or to give suggestions for improvement.

It will speed up certain processes and allow focus on the actual thinking and more enjoyable parts. As of now, AI is too generic to be truly helpful, but I use it to come up with different ideas and get a first overview.

As it is currently, we take everything with a grain of salt, so I am not sure how reliable the outcomes are. A sense check is always required for even the lowest-level output. I think it could be a great complementary tool.

**Table B.1 – continued from previous page**

---

**What do you think about integrating Generative AI in strategic planning activities? (continued)**

---

It is very useful for a brainstorming phase, but the more you use it for generating full analyses, the more of a black box it becomes, making it less transparent and harder to justify as a consultant.

It can help as a starting point but needs adjustments to make it tangible and client-specific.

I like it, but it should keep thinking outside the box and also provide totally new opportunities. As long as it does not do that, it cannot replace people.

Using Generative AI in strategic planning is like having a supercharged brainstorming partner—it helps explore ideas, spot patterns and test scenarios faster. It does not replace human judgment, but it makes the process smarter and more dynamic.

Generative AI can be used well for the initial design of strategic planning activities, but humans will always have to make additions and checks. It is also important that the right information is given (right prompts) and that follow-up questions are asked.

---

Table B.2: Respondents' examples of situations where Generative AI supported strategic planning activities and what worked well.

---

**Can you describe a situation, if any, where Generative AI supported you in a strategic planning activity as mentioned in the survey? What worked well?**

---

Risk identification – GenAI is good at providing high-level risks in different areas, some of which can be easily overlooked.

Several situations: risk analysis, stakeholder analysis and Five Forces.

During our SAP implementation, we used Generative AI to simulate different rollout strategies and flag potential integration issues before they became real problems. What really stood out was how it helped us align stakeholders quickly by turning complex data into clear, actionable insights.

It was able to identify risks that were unknown to me – so-called unknown unknowns. That is great for raising awareness of risks.

Using it for a SWOT-style activity, where Generative AI helped me think about strengths, weaknesses, opportunities and threats for the company. I did find some errors, though, highlighting the importance of double-checking.

SWOT and risk assessments – definitely works well.

I used Copilot for SWOT evaluation. With a good prompt and clear context (without disclosing the client), the results were impressive. After several iterations and refinements, the output was great.

Designing the framework for an analysis. This worked very well and the client was happy.

---

**Table B.2 – continued from previous page**

**Can you describe a situation, if any, where Generative AI supported you in a strategic planning activity as mentioned in the survey? What worked well? (continued)**

Mostly research and summarizing, but also identifying risks or helping translate project scopes into business value and making this more quantifiable.

Doing background research on an industry in a couple of minutes, saving lots of time.

It saved me time when diving deep into a sector and it also came up with out-of-the-box ideas.

I use GenAI on a daily basis for maturity assessments as well as identifying market and technology trends.

Reviewing the options on the table and using it almost as a sparring partner to brainstorm with before human discussions.

Often brainstorming on risks or improvement possibilities for a client when I describe my situation to GenAI.

Table B.3: Respondents' reported limitations and risks when using Generative AI for strategic-planning activities.

**Describe one limitation or risk you experienced when using Generative AI for strategic-planning activities.**

A tendency to generate overly confident recommendations based on incomplete or outdated data.

Overly convincing answers can make you lazy to read critically; outputs are not always the right solution or formulation.

Hallucinations are a key risk; effective guardrails are needed.

Undetectable hallucinations; convergence risk—if multiple organizations use the same model, inputs become similar and competitive advantage erodes.

Conformity/authority bias: “ready-to-go” outputs may be accepted too quickly; risk of groupthink (people align with perceived authority even if they disagree).

Enterprise setup does not support the most recent LLMs.

Risk of overreliance—belief that AI will outperform humans in 2–3 years could reduce human oversight.

Outsourcing one's thinking: plausible but incorrect answers can lead to insufficient analysis.

“Garbage in, garbage out”; input and output quality assurance is required.

Company LLMs generally provide less accurate results than LLM+RAG setups built into platforms.

Weak problem understanding without explicit framing; tends toward generic analysis.

Outputs are too generic; risk of overlap across engagements.

Erosion of critical thinking and attention; decisions may become more fragile and superficial.

**Table B.3 – continued from previous page**

**Describe one limitation or risk you experienced when using Generative AI for strategic-planning activities. (continued)**

Internal AiQ engine is limited; desire to load frameworks and client context to generate logical strategic directions.

Limited reasoning and context window; hallucinations persist—larger context/project memory would help.

Copy–paste risk without proper verification.

Hard to leverage internal documents/knowledge (e.g., KPMG inputs) to enhance results.

Tailoring requires extensive client data that is currently infeasible to provide; otherwise answers remain generic.

Validation burden: with large inputs (multiple files, studies), users may be unable to verify and may accept outputs as truth.

You must remain critical in your thinking.

Hallucinated, impossible scenarios.

Data confidentiality concerns.

Lacks deep understanding of the company and its people (tacit context).

Table B.4: Consultants’ suggestions for improving the effective use of Generative AI in strategic planning.

**What kind of training, tools or support would help you use Generative AI more effectively in strategic planning?**

Training or guidelines on how to evaluate AI output

Training in Agents

Limitations and risks are more relevant compared to smart prompting

Introduction to AI

Clear showcase of value (e.g., normally deliverable A costs 40 hours work, now 4)

Business support

Access to latest models. Current enterprise approach limits access to reasoning models or building chains of analysis with different LLMs

Table B.5: Consultants’ self-reported prompting practices when using Generative AI tools.

**Which of the following best describes how you use prompts when working with Generative AI tools?**

Use prompts colleagues sent me

**Table B.5 – continued from previous page**

---

**Which of the following best describes how you use prompts when working with Generative AI tools? (continued)**

---

I use LLMs to write the prompt, taking into account research on best practices and emerging prompting techniques

I typically ask for source information in a tedious way using prompting

Combination of the above

---

# Chapter C Practical demonstration

This appendix document translates the study's empirical insights into practice by demonstrating how generative AI can support priority strategic planning activities in a management consulting context identified in Chapter 5. It completes the exploratory sequential design by adding an applied component to the qualitative and quantitative phases and shows what this looks like with reproducible artifacts. Full prompts and long-form outputs are included in the Appendix D, for transparency and replication. This practical demonstration uses ChatGPT-5. To ensure a fair comparison between prompting strategies, the model configuration is held constant across all activities.

The selection of activities is grounded in the thesis' findings. The Discussion (Chapter 5) highlights one clear success story for generative AI: the structured analytical frameworks, such as SWOT, PESTLE and Porter, where satisfaction and importance are relatively high and gaps are smaller. For the demonstration, SWOT is chosen as a representative activity from the mature cluster. SWOT provides a stable and familiar framework that fits well with the strengths of LLMs, since it is rule-based and can be broken into clear categories. Focusing on one activity rather than many allows for more depth, showing in detail how prompt design can improve the quality, structure and actionability of outputs.

The purpose of this chapter is to enrich the quantitative findings by providing examples of generative AI usage in practice on strategic planning activities. By showcasing the outputs of the LLM it becomes increasingly clear what its potential is. To realize this, prompting is approached as an iterative design process, where simple starting points are progressively refined through prompt engineering. The prompting strategies discussed in Section 2.1.4.2 are zero-shot, few-shot, Chain-of-Thought and Retrieval-Augmentation Generation and are not treated as disjoint techniques but as composable building blocks that can be layered where they add the most value. Each has particular strengths and weaknesses, as summarized in Table 2.1. Besides, these various prompting strategies may be combined to achieve even better prompting results. Prompting strategies can also be enhanced by providing, role, specification, structure guidance, exemplars, reasoning instructions and retrieval options. The demonstration therefore follows a staged refinement process, beginning with a baseline SWOT prompt and then incrementally adding role specifications, exemplar, reasoning steps or grounding elements to illustrate how deliberate prompt design can shift outputs from generic to consultant-ready.

The demonstration will follow a fixed protocol to keep the comparison auditable and focused on the effects of prompt design:

1. **Task framing:** the business context is clearly explained and its goal.

2. **Baseline SWOT:** a minimal zero-shot prompt aligned to SWOT's four categories, including 2-3 "strategic implications".
3. **Refined prompt:** re-run the SWOT and implications with an explicit consultant role, audience, task, style and structure guidance to increase relevance and completeness.
4. **Reasoning overlay:** add reasoning instruction to improve formatting, depth and transparency, by requiring mapping drivers into S/W/O/T before composing implications.
5. **Few-shot micro exemplars:** provide two short exemplars that demonstrate desired tone, depth and formatting, then apply the same structure to the common case to lock form and specificity.
6. **Grounded variant:** provide 4 curated public text snippets (e.g., news reports, analyst notes, reports) each with source tag [S#] and require inline attributions on any evidence-based bullet.
7. **Targeted refinement & finalize:** prompt a brief self-critique against four lightweight criteria (structure adherence, quadrant completeness, clarity of implications, explicit assumptions) and revise against the rubric, where confidence is low and flag any remaining uncertainties once to produce the final illustrative artifact.

For each variant, the prompt, the model's output and a two-to-three sentence design note explaining what changed and why are archived in the Appendix D. This section will only contain concise excerpts to illustrate differences that matter for consulting practice.

Although strategic implications are not part of the canonical SWOT, they are included here to demonstrate how generative AI can bridge from analysis to action, which is standard practice in consulting deliverables. In this context, strategic implications refer to short, forward-looking statements that connect internal factors (strengths and weaknesses) with external conditions (opportunities and threats) to suggest a possible course of action. For example, a strength in strong client relationships combined with an opportunity in growing market demand may imply prioritizing partnership-based expansion. Including such implications makes the analysis more actionable and illustrates how LLMs can move beyond categorization toward informing strategic decision-making. To maintain focus on prompt design rather than on factual correctness, the demonstration is based on a synthetic case scenario: a European HR-tech scale-up AI-based talent matching platform. Using a fictive case avoids distracting debates about company-specific accuracy while still providing a realistic context. This common scenario is presented at the start of each prompt and remains unchanged across all stages:

*A European HR-tech scale-up is developing a generative AI service that helps enterprises with workforce planning. The platform uses LLMs to scan job descriptions, internal performance data and training catalogs to generate skill-gap analyses and propose reskilling strategies. The service targets organizations facing talent shortages in digital and engineering roles.*

## C.1 Prompt engineering

This section demonstrates how deliberate prompt engineering shapes the performance of generative AI in strategic planning tasks, specifically the SWOT analysis. To ensure comparability, each prompt was entered into a blank prompting page so that no memory of earlier runs could influence later ones. This design reduces bias and makes it possible to compare outputs side by side. Only the final stage (Stage 6) was prompted in the same prompting page, as it specifically relies on the model critiquing and improving its own draft.

The process follows six stages, which are all distinct and incrementally refined per stage. Each stage is documented with full input and output in Appendix D. For each stage, the changes in prompt design and their effects are visible, allowing the reader to see how simple adjustments can shift outputs. First, the stages are mentioned briefly below, at the end an overview is presented in Table C.1.

**Stage 1 - Baseline SWOT** The first stage uses a minimal zero-shot prompt (see Figure D.1). The output resembles raw brainstorming notes: broad ideas are captured, but categories overlap and implications remain vague. The value for consultants lies in seeing the "default" starting point of an unstructured model. The drawback is clear: without design, outputs remain too generic to be of practical use.

**Stage 2 - Refined prompt** Adding role framing, style and structure guidance (see Figures D.2 and D.3) immediately improves organization. The SWOT becomes more balanced and implications are sharper. What stands out is how small changes in wording make the output look somewhat closer to consulting deliverables.

**Stage 3 - Reasoning overlay** Here, the model is asked to first list drivers before mapping them into SWOT quadrants (see Figures D.4 to D.6). This adds transparency, as the logic behind each item mapping becomes clearer. The benefit is that it reduces overlap and shows why each point sits where it does. What stands out is the clearer link between drivers and the final SWOT, whereas in previous stages the SWOT was just produced, without backed information.

**Stage 4 - Few-shot exemplars** By including an exemplar SWOT (see Figures D.7 to D.9), the model follows the exact structure of this exemplar, providing clear guidance. The output looks more polished and consistent, with a tone that resembles consulting slides.

**Stage 5 - Grounded variant** In this stage, public external sources are introduced and referenced to with [S] (see Figures D.10 to D.13). This makes the SWOT evidence-based and auditable, introducing stronger claims. The value is that claims are not traceable to real sources, as consultants would do in practice. What stands out is the explicit citation of evidence, moving the analysis from plausible to credible.

**Stage 6 - Targeted refinement & finalization** Finally, the model critiques its own Stage 5 output against a rubric (structure, completeness, clarity, assumptions) and refines it once (see Figures D.14 to D.17). This produces an even more focused, streamlined output with assumptions called out and criticized by itself. The benefit is showing generative AI not just as

a generator, but also as a reviewer. What stands out is the shift toward a "final draft" quality, which is now ready for expert consultant review.

Table C.1: Overview of prompting stages, benefits and limitations

Stage	Prompting Strategy	Change Introduced	Benefit	Limitation
1	Baseline SWOT (Zero-shot)	Minimal prompt with no structure or role	Neutral benchmark, illustrates unrefined output	Generic, overlapping categories, weak implications
2	Refined prompt	Role framing, audience and structural guidance	More complete organized, aligned to consulting deliverables	Reasoning behind choices remains opaque
3	Reasoning overlay (CoT)	Drivers listed first, then mapped into SWOT before implications	Transparent causal chain, reduced drift, auditability	Longer prompts, higher token cost
4	Few-shot exemplars	A short exemplar SWOTs lock tone, style, depth	Mimics consulting-grade outputs, consistency across cases	Exemplar quality critical; poor examples propagate errors
5	Grounded variant (Simulated RAG)	Curated public sources with inline attributions [S#]	Evidence-based, auditable outputs closer to practice	Dependent on source selection, potential bias
6	Targeted refinement	Self-critique against rubric, one-shot revision	Iterative improvement, AI as reviewer, higher clarity	Limited domain validation, may miss blind spots

Finally, the six stages illustrate the trajectory from raw brainstorming to client-ready deliverables. Each stage mirrors a familiar consulting workflow: Stage 1 resembles brainstorming notes on a whiteboard, Stage 2 the first outline slides, Stage 3 the driver trees and causal logic consultants use to justify their framing, Stage 4 the application of a consistent "house style", Stage 5 the evidence-backed deck enriched with external data and Stage 6 the final quality-control pass by a senior consultant or partner (AI in this case). For consultants, the implication is clear: mastering prompt engineering allows generative AI to be integrated not just as a productivity tool, but as a strategic sparring partner across the full lifecycle of strategy deliverable preparation.

# Chapter D Output of LLMs

This section presents the exact input prompts and outputs generated by the LLM for the practical demonstration. The process consists of six stages. Each stage shows the full prompt text and corresponding model output. In addition, a short design note of 2–3 sentences explains what changed in the prompt design compared to the previous stage and what effect this had on the output. For Stage 1 (Baseline SWOT), the change is not included.

## D.1 Stage 1 – Baseline SWOT

This stage uses a minimal zero-shot prompt. The output provides a basic SWOT analysis with short implications but tends to be broad and generic, serving as a reference point for further refinements.

Figure D.1: Stage 1 – Zero-shot SWOT with broad generic items and overlap.

A European HR-tech scale-up is developing a generative AI service that helps enterprises with workforce planning. The platform uses LLMs to scan job descriptions, internal performance data and training catalogs to generate skill-gap analyses and propose reskilling strategies. The service targets organizations facing talent shortages in digital and engineering roles.

Please provide a SWOT analysis for this company based on the scenario described. After the SWOT, add 2–3 short strategic implications that explain what the factors mean for the company's choices.

Here's a structured SWOT analysis for the HR-tech scale-up, followed by concise strategic implications:

### SWOT Analysis

#### Strengths

- **Innovative value proposition:** Unique use of generative AI for skill-gap analysis and reskilling, directly addressing a major pain point.
- **High enterprise relevance:** Targets workforce planning in industries with acute digital and engineering talent shortages.
- **Data leverage:** Integrates multiple data sources (job descriptions, performance data, training catalogs), enabling richer insights than traditional HR tools.
- **Scalability:** Cloud-based generative AI services can expand quickly across markets.

#### Weaknesses

- **Data dependency:** Effectiveness relies on quality and accessibility of client data, which may be fragmented or incomplete.
- **Trust and adoption barriers:** Enterprises may be cautious about sensitive HR data use, model transparency, and bias risks.
- **Limited track record:** As a scale-up, credibility and proven ROI may be questioned compared to established HR software vendors.
- **High compute costs:** Running LLMs at scale can strain margins if not optimized.

#### Opportunities

- **Growing skills gap:** Enterprises are under pressure to reskill workforces in digital/engineering fields; demand is rising.
- **AI adoption wave in HR:** Organizations are actively seeking AI-driven HR tools, opening a receptive market.
- **Partnerships with training providers:** Can integrate with learning platforms to offer end-to-end workforce development solutions.
- **Expansion beyond digital/engineering:** The model could extend to other domains (healthcare, green tech, etc.) facing shortages.

#### Threats

- **Competitive landscape:** Established HR platforms (Workday, SAP SuccessFactors, LinkedIn Learning) could replicate features.
- **Regulatory risks:** EU AI Act and data privacy regulations may impose constraints on HR data processing and model explainability.
- **Talent war in AI:** Recruiting and retaining AI/ML talent for continued product innovation is costly and competitive.
- **Enterprise inertia:** Large organizations may resist change or take years to adopt new HR tech solutions.

### Strategic Implications

1. **Differentiate through trust and compliance:** Building explainable AI features and strong GDPR/AI Act compliance can turn a weakness (data sensitivity) into a competitive edge.
2. **Form ecosystem partnerships:** Collaborations with training providers and HRIS vendors can enhance value and accelerate adoption.
3. **Focus initial growth on high-shortage sectors:** Prioritize digital and engineering domains to establish market leadership before expanding to adjacent industries.

## **D.2 Stage 2 – Refined prompt**

This stage introduces an explicit consultant role, sector context and structural guidance to increase the specificity and completeness of the SWOT. The output is more organized and provides more context-aware output.

Figure D.2: Stage 2 – Role and strict schema make items concise and consistent, part 1

You are a strategy consultant preparing a client-ready note for senior executives.

**Context:**  
A European HR-tech scale-up is developing a generative AI service that helps enterprises with workforce planning. The platform uses LLMs to scan job descriptions, internal performance data, and training catalogs to generate skill-gap analyses and propose reskilling strategies. The service targets organizations facing talent shortages in digital and engineering roles.

**Task:**  
Prepare a structured SWOT analysis of this company based on the scenario above.

**Output format:**  
A) Strengths (internal) – up to 6 concise bullets  
B) Weaknesses (internal) – up to 6 concise bullets  
C) Opportunities (external) – up to 6 concise bullets  
D) Threats (external) – up to 6 concise bullets  
E) Strategic implications – 2–3 bullets linking internal factors (S/W) with external factors (O/T), phrased as short recommendations for management

**Style:**  
- Professional consulting tone, American English  
- Clear, crisp bullet points  
- No generic phrases like “innovative” or “fast growing” without specifying why

Here's a client-ready note in the requested structure and tone:

#### **A) Strengths (internal)**

- Proprietary use of LLMs tailored to workforce planning, creating differentiated skill-gap insights
- Ability to integrate multiple data sources (job descriptions, performance data, training catalogs) into actionable recommendations
- Strong focus on digital and engineering roles, aligning with critical shortages across Europe
- Scalable SaaS delivery model enabling cross-industry applicability
- Potential for recurring revenue streams through subscription and enterprise licensing

#### **B) Weaknesses (internal)**

- Heavy reliance on quality and consistency of client-provided data; poor inputs may limit output accuracy
- Limited track record in large-scale enterprise deployments may reduce credibility with risk-averse HR leaders
- Significant compute costs tied to LLM use, challenging unit economics at scale
- Narrow domain scope (digital and engineering roles) may restrict adoption beyond early target segments
- Possible internal capability gaps in enterprise sales and customer success functions

#### **C) Opportunities (external)**

- Widespread labor shortages in STEM and digital fields driving demand for advanced workforce planning tools
- Growing corporate investment in reskilling and upskilling programs supported by EU and national funding initiatives
- Increasing acceptance of AI-driven decision support in HR processes
- Potential to expand into adjacent use cases (e.g., succession planning, internal mobility, diversity analytics)
- Partnerships with training providers, universities, or LMS vendors to strengthen value proposition

Figure D.3: Stage 2 – Role and strict schema make items concise and consistent, part 2

**D) Threats (external)**

- Intensifying competition from large HR-tech incumbents (e.g., Workday, SAP SuccessFactors) integrating AI into their platforms
- Regulatory scrutiny over algorithmic bias, data privacy, and transparency in HR AI applications
- Client hesitancy to trust generative AI outputs for sensitive workforce decisions
- Economic downturns could deprioritize HR innovation budgets despite skill shortages
- Rapid pace of AI model commoditization eroding differentiation over time

**E) Strategic implications**

- **Build enterprise credibility:** Mitigate weaknesses in track record by pursuing pilots with flagship clients and publishing validated case studies to strengthen positioning against incumbents.
- **Anchor in reskilling ecosystem:** Leverage strengths in skills-gap analytics to form partnerships with training providers, aligning internal capabilities with external funding and demand.
- **Control risk perception:** Address threats of bias and trust by embedding transparency features and compliance-by-design, turning regulatory scrutiny into a competitive advantage.

---

Would you like me to also visualize this SWOT in a 2x2 matrix format (consulting-style slide) so you can use it directly in a deck?

### **D.3 Stage 3 – Reasoning overlay (Chain-of-Thought)**

Here the prompt is extended with reasoning instructions. The model is required to first list drivers before mapping them into SWOT categories. This makes the reasoning transparent and helps reduce off-topic or repetitive factors.

Figure D.4: Stage 3 – Examples guide tone and raise specificity for this domain, part 1

You are a strategy consultant preparing a client-ready note for senior executives.

**Context:**  
A European HR-tech scale-up is developing a generative AI service that helps enterprises with workforce planning. The platform uses LLMs to scan job descriptions, internal performance data, and training catalogs to generate skill-gap analyses and propose reskilling strategies. The service targets organizations facing talent shortages in digital and engineering roles.

**Task:**  
You should think step by step.  
Step 1: Identify 5-8 underlying drivers (mix of internal and external)  
Step 2: For each driver, classify it into S, W, O or T and explain your reasoning in 1-2 sentences.  
Step 3: Produce a clean SWOT with concise bullets under each heading.  
Step 4: Add 2-3 strategic implications that link S/W with O/T, make sure to provide your reasoning.

**Output format:**  
1) Key drivers (5-8 concise bullets, mix of internal and external, including 1-2 sentences explanation)  
2) SWOT with clear headings:  
A) Strengths (internal) – up to 6 concise bullets  
B) Weaknesses (internal) – up to 6 concise bullets  
C) Opportunities (external) – up to 6 concise bullets  
D) Threats (external) – up to 6 concise bullets  
3) Strategic implications – 2-3 short bullets linking S/W with O/T, phrased as directional recommendations

**Style:**  
- Professional consulting tone, American English  
- Clear, crisp bullet points  
- Each driver must logically connect to at least one SWOT item  
- No generic statements like "innovative company"; always specify context

Nagedacht

Here's a tight, executive-ready build that walks through the steps and keeps every driver tied to the SWOT.

### 1) Key drivers (with classification and brief reasoning)

- **Acute shortages in digital and engineering roles [O]**  
Persistent hiring gaps and rising skill requirements increase demand for solutions that quantify skill gaps and enable reskilling versus hiring.
- **Enterprise data landscape: job descriptions, performance data, and training catalogs [S]**  
The product natively ingests these sources, enabling a holistic "skills supply vs demand" view that point tools or manual analysis cannot match.
- **Model quality and explainability for skill inference [W/T]**  
LLMs can hallucinate or misclassify skills; without transparent mappings and audit trails, HR and Risk functions will block adoption.
- **Integration depth with HRIS/LMS and workflow tooling [S/W]**  
Where connectors and write-backs exist, the product becomes embedded in daily HR workflows; gaps in integrations slow time-to-value.
- **Proof of value and ROI quantification [W]**  
Buyers need clear, defensible outcomes (e.g., reduced time-to-fill, lower training spend, internal mobility uplift). Early-stage metrics may be thin.
- **EU regulatory pressure (AI Act, GDPR) on HR data processing [T/O]**  
Compliance demands can slow sales cycles, yet vendors with strong governance, data minimization, and auditability can differentiate.

Figure D.5: Stage 3 – Examples guide tone and raise specificity for this domain, part 2

- **Incumbent HCM suites bundling AI features [T]**  
Workday/SAP/Oracle can ship “good enough” skill-gap features at marginal cost, pressuring pricing and upsell.
  - **Shift from external hiring to internal mobility and reskilling [O]**  
Budget-constrained HR leaders favor redeploying talent; solutions that map skills to training pathways align directly with this pivot.
- 

## 2) SWOT

### A) Strengths (internal)

- End-to-end view: combines job demand, employee performance, and training supply to produce actionable skill-gap analyses.
- LLM-driven skill extraction that detects emerging tech skills earlier than static taxonomies.
- Growing integration footprint across HRIS/LMS increases embedment in core HR workflows.
- Clear reskilling workflows that translate gaps into training paths, not just dashboards.
- Focus on digital/engineering roles creates sharper models and faster early wins.

### B) Weaknesses (internal)

- Limited explainability and lineage from text → skills → recommendations undermines trust and adoption.
- Evidence of ROI may be nascent, with few longitudinal case studies.
- Dependence on customer data quality; messy job titles and sparse performance data degrade outputs.
- Integration gaps or brittle connectors delay implementation and value realization.
- Human-in-the-loop review process may not yet scale for large enterprises.

### C) Opportunities (external)

- Worsening digital talent shortages drive demand for reskilling over net-new hiring.
- Internal mobility and redeployment agendas need skill visibility and training alignment.
- Regulatory push for governance lets a transparent, auditable solution stand out.
- Partnerships with training providers to auto-build role-based learning paths.
- Expansion from engineering into adjacent domains (e.g., data, product, cyber).
- Privacy-preserving benchmarking (with consent) to show relative skill maturity versus peers.

### D) Threats (external)

- HCM/ERP incumbents bundle “good enough” AI skill features, compressing price and differentiation.
  - AI Act/GDPR enforcement slows or restricts LLM use on HR data without strong controls.
  - Shadow AI or generic LLM tools tempt analysts to DIY, reducing perceived need to procure.
  - Prolonged budget scrutiny elongates enterprise sales cycles and pushes pilots over scale.
  - Data security incidents (sector-wide or vendor-specific) erode trust quickly in HR contexts.
-

Figure D.6: Stage 3 – Examples guide tone and raise specificity for this domain, part 3

### 3) Strategic implications (linking S/W with O/T)

- **Make explainability the moat to convert regulation into advantage.**  
Address **W: explainability** and **W: ROI evidence** against **T: AI Act/GDPR** by shipping skill lineage (source text → extracted skill → confidence), human-in-the-loop approvals, and audit logs. This turns **O: regulatory differentiation** into a clear win theme and de-risks enterprise sign-off.
- **Lock in via integrations and training-pathway partnerships to withstand HCM bundling.**  
Use **S: end-to-end data view** and **S: workflow integrations** to become the orchestration layer that incumbents can't easily replace, then deepen with **O: training partnerships** that auto-generate role paths. This blunts **T: incumbent bundling** and accelerates time-to-value.
- **Productize a "Data Readiness + ROI" playbook to speed decisions and scale proof.**  
Tackle **W: data quality** and **W: ROI** by offering a fixed-scope data normalization accelerator, baseline skill inventory, and a 90-day ROI template (time-to-fill reduction, redeployment rate, training cost per uplift). This aligns with **O: reskilling/internal mobility** while reducing **T: budget scrutiny** and pilot purgatory.

If you want, I can turn this into a one-pager slide with a headline, the SWOT, and a "What we recommend" box.

## **D.4 Stage 4 – Few-shot micro exemplars**

A short exemplar SWOT is added to the prompt. These lock in the tone, structure and granularity, leading to outputs that more closely mirror the style and depth of consulting deliverables.

Figure D.7: Stage 4 – Stepwise prompting clarifies logic and reduces quadrant leakage, part 1

You are a strategy consultant preparing a client-ready note for senior executives.

**Context:**  
A European HR-tech scale-up is developing a generative AI service that helps enterprises with workforce planning. The platform uses LLMs to scan job descriptions, internal performance data, and training catalogs to generate skill-gap analyses and propose reskilling strategies. The service targets organizations facing talent shortages in digital and engineering roles.

**Task:**  
Prepare a structured SWOT analysis for the scenario above.  
Follow the exemplar's format exactly: list Key Drivers (with IDs), then produce SWOT bullets that include a short rationale and explicit links back to driver IDs, and finish with 2–3 strategic implications that tie S/W to O/T.

**Bullet schema (use this micro-format for every SWOT item):**  
• <Claim> — Why it matters: <short rationale> — Link: [D#,...]

**EXEMPLAR (use only to learn structure, depth, tone; do NOT copy content):**

**Case:** European mid-market SaaS provider for HR analytics

**Key Drivers (assign IDs D1...):**  
D1. Enterprises report persistent shortages in digital/engineering roles across EU.  
D2. Procurement scrutiny rises for explainable/ethical AI (EU AI Act pressure).  
D3. HR suite vendors embed generative features, creating bundle pricing pressure.  
D4. In-house LLM/fine-tuning talent enables fast iteration on client requests.  
D5. HR data is fragmented across HRIS/LMS; integration quality varies.  
D6. Training providers seek partnerships to bundle skills pathways for enterprises.

**SWOT (use the schema; each bullet must cite ≥1 driver ID):**

**Strengths**  
• Unified data-integration pipeline across HRIS/LMS — Why it matters: enables deeper, cross-source skill-gap insights — Link: [D5,D4]

**Weaknesses**  
• Limited enterprise sales coverage and procurement navigation — Why it matters: slows large-deal conversion under AI governance checks — Link: [D2]

**Opportunities**  
• Growing demand for bundled "assess→train→track" reskilling pathways — Why it matters: opens end-to-end offers with higher ROI clarity — Link: [D1,D6]

**Threats**  
• Suite vendors bundle AI features at zero incremental price — Why it matters: compresses willingness to pay for stand-alone tools — Link: [D3]

**Strategic Implications (tie S/W ↔ O/T explicitly):**  
• Leverage integration strength (S) to co-build pathways with training partners (O) and defend value versus suite bundling (T).  
• Address enterprise sales/procurement gap (W) to pass AI-governance gates (T) and unlock deals in shortage-heavy segments (O).

Figure D.8: Stage 4 – Stepwise prompting clarifies logic and reduces quadrant leakage, part 2

— End of exemplar —

Now apply the same structure and tone to the HR-tech scale-up case.

Output format:

- 1) Key drivers (5–8 concise bullets, mix of internal and external)
- 2) SWOT with clear headings:
  - A) Strengths – up to 6 bullets, each specific and non-overlapping
  - B) Weaknesses – up to 6 bullets, each specific and non-overlapping
  - C) Opportunities – up to 6 bullets, each specific and non-overlapping
  - D) Threats – up to 6 bullets, each specific and non-overlapping
- 3) Strategic implications – 2–3 bullets, each explicitly linking an internal factor (S/W) with an external factor (O/T) in a cause–effect phrasing (e.g., "Leverage X (S) because Y (O)" or "Address Z (W) to mitigate A (T)").

Style:

- Professional consulting tone, American English
- Crisp, precise, sector-specific bullets
- Avoid vague wording like "innovative" or "fast-growing"; always specify context
- Follow the structure and depth modeled in the exemplar

Here's the structured, client-ready note in the exact format you asked for:

---

**Case: European HR-tech scale-up – Generative AI workforce planning platform**

---

**1) Key Drivers**

- D1. Persistent EU-wide shortages in digital and engineering talent raise demand for workforce reskilling solutions.
- D2. EU AI Act and corporate governance frameworks increase scrutiny of explainability and ethical AI use.
- D3. Large HR-suite vendors (Workday, SAP, Oracle) are embedding generative AI features, raising bundling pressure.
- D4. The scale-up has proprietary LLM fine-tuning capabilities for HR language and skill taxonomies.
- D5. Client HR data is dispersed across HRIS, LMS, and performance systems, making integration complex.
- D6. Training providers seek tech partners to connect assessment with curated learning paths.
- D7. Mid-market enterprises lack in-house AI expertise, creating reliance on external platforms.

---

**2) SWOT**

**Strengths**

- Proprietary fine-tuning on HR and skills taxonomies — Why it matters: improves accuracy of skill-gap detection versus generic LLMs — Link: [D4]
- Early credibility in addressing fragmented HR data — Why it matters: proven ability to harmonize inputs across HRIS/LMS/performance systems — Link: [D5]

**Weaknesses**

- Limited procurement track record with large enterprises — Why it matters: slows deal cycles under stricter AI governance checks — Link: [D2]
- Narrow go-to-market capacity — Why it matters: constrains reach in a fast-moving, bundle-dominated HR-tech market — Link: [D3]

**Opportunities**

- Rising demand for reskilling pathways in digital/engineering shortages — Why it matters: creates willingness to invest in ROI-linked solutions — Link: [D1,D7]

## Figure D.9: Stage 4 – Stepwise prompting clarifies logic and reduces quadrant leakage, part 3

- Partnering with training providers to offer end-to-end “assess→train→track” journeys  
– Why it matters: enables defensible differentiation beyond analytics — Link: [D6]

### Threats

- HR-suite vendors embedding AI features at no incremental cost — Why it matters: erodes standalone value perception — Link: [D3]
- Stricter explainability and compliance requirements — Why it matters: raises bar for SMEs without established governance processes — Link: [D2]

---

### 3) Strategic Implications

- Leverage proprietary fine-tuning strength (S) to co-develop bundled reskilling pathways with training providers (O), countering suite vendor bundling pressure (T).
- Address enterprise procurement and governance weakness (W) to satisfy AI-compliance checks (T) and unlock adoption in shortage-driven sectors (O).
- Expand go-to-market reach (W) by targeting mid-market firms reliant on external AI expertise (O), where vendor bundling threat (T) is less pronounced.

---

Would you like me to **extend each quadrant to 4–6 bullets** for extra depth (matching the exemplar’s “up to 6 per category”), or keep it lean at 2 per quadrant for maximum crispness?

## **D.5 Stage 5 – Grounded variant (Simulated RAG)**

This stage asks the model to browse publicly available renowned sources. The aim is to simulate retrieval-augmented generation, improving grounding and auditability of the SWOT factors.

Figure D.10: Stage 5 – Grounded in source snippets so claims are evidence based, part 1

You are a strategy consultant preparing a client-ready note for senior executives.

**Context:**  
A European HR-tech scale-up is developing a generative AI service that helps enterprises with workforce planning. The platform uses LLMs to scan job descriptions, internal performance data, and training catalogs to generate skill-gap analyses and propose reskilling strategies. The service targets organizations facing talent shortages in digital and engineering roles.

**Task:**  
Prepare a structured SWOT analysis for this company based on the scenario above. Ground your analysis in publicly available external sources. Every claim based on external evidence must be tagged with a citation [S#]. At the end, provide a "Source List" with [S#, starting at S1, S2, S3 etc.] mapped to the full reference (title, author/organization, year, and URL).

**Examples of relevant sources (you may select from or extend beyond these):**

- Market & Trend Data (The "Why")**
  - European Commission / Eurostat (DESI Index) – data on EU digital skills gap.
  - World Economic Forum (Future of Jobs Report) – roles in demand, reskilling pace.
  - McKinsey Global Institute (MGI) / BCG – impact of automation, enterprise upskilling.
- Technology & Methodology (The "How")**
  - Academic databases (Google Scholar / ResearchGate) – papers on GenAI for HR, skill-gap modeling.
  - Open-source LLM documentation (Hugging Face, GitHub datasets) – evidence of technical feasibility and bias.
  - Analyst firms (Gartner / Forrester) – public reports on HCM/HR-tech market trends.
- Competitive Landscape & Events (The "Who")**
  - HR-tech conferences (UNLEASH, Web Summit, TechCrunch Disrupt) – competitor launches, ecosystem trends.
  - Venture capital databases/news (Crunchbase, Sifted) – European HR-tech funding, acquisitions.
  - Company blogs / LinkedIn posts – messaging by peer HR-tech firms.

**Process:**

- 1) Identify 5–7 relevant public sources and summarize them briefly as "Key Sources" at the start.
- 2) Use these sources to build Key Drivers, SWOT factors, and Strategic Implications.
- 3) Each SWOT bullet must include at least one inline driver link [D#] and, if evidence-based, a source tag [S#].
- 4) At the end, list all sources in a "Source List."

**Output format:**

- 1) Key Sources — 3–5 short summaries with [S#].
- 2) Key Drivers — 5–8 concise drivers (internal + external), link to sources [S#].
- 3) SWOT (each bullet must follow schema: Claim — Why it matters — Link: [D#,...] [S#?]):
  - A) Strengths (internal)
  - B) Weaknesses (internal)
  - C) Opportunities (external)
  - D) Threats (external)
- 4) Strategic Implications — 2–3 bullets explicitly linking S/W ↔ O/T, citing [S#] where relevant.
- 5) Source List — map [S#] to title, author/organization, year, and URL.

Figure D.11: Stage 5 – Grounded in source snippets so claims are evidence based, part 2

Style & constraints:

- Professional consulting tone, American English.
- Only use publicly available sources.
- Always include inline [S#] attribution where evidence is used.
- If no evidence exists, write "Unknown."
- Avoid overlap across SWOT categories.

📄 🔗 < 2/2 >

Nagedacht

Here's a crisp, source-grounded SWOT you can drop into a client note.

### 1) Key Sources

- [S1] EU "State of the Digital Decade 2024" — Only 55.6% of EU citizens have at least basic digital skills; ICT specialists projected to reach ~12M by 2030 vs. 20M target. Evidence of persistent EU skills gap and policy urgency. [S1]
- [S2] WEF *Future of Jobs 2025* — Employers expect ~39% of core skills to change by 2030; strong emphasis on continuous learning and reskilling. Quantifies skilling velocity and areas in demand. [S2]
- [S3] McKinsey Global Institute (2024) — AI/automation shift demand toward STEM and high-skill roles; companies plan to emphasize retraining alongside hiring/subcontracting. Validates reskilling demand from enterprises. [S3]
- [S4] EU AI Act — Entered into force Aug 1, 2024; GPAI obligations apply from Aug 2, 2025; broader high-risk rules phased in by 2026–2027. Confirms rising compliance bar for AI in HR contexts. [S4]
- [S5] arXiv (2025): Cultural bias in LLM-based hiring evaluations — Documents scoring disparities and bias risks for LLMs in hiring-adjacent use cases. Underscores need for fairness controls in HR applications. [S5]
- [S6] Gartner (public page) — HR leaders prioritizing AI use cases tied to skills, with emphasis on accuracy, reliability, transparency, and partnering with IT/legal/compliance. Signals buyer priorities and evaluation criteria. [S6]
- [S7] Crunchbase News (2025) — HR software funding: ~\$2B in 2024 (well below 2021 peak), with modest uptick in 2025 from a low base. Funding conditions remain selective. Informs go-to-market and runway assumptions. [S7]

### 2) Key Drivers

- D1. Structural EU digital skills gap persists** — EU far from 2030 targets; <60% with basic digital skills and ICT specialists below target trajectory. [S1]
- D2. Rapid skill change & reskilling imperative** — ~39% of core skills change by 2030; enterprises leaning into continuous learning. [S2][S3]
- D3. Regulatory pressure (EU AI Act)** — GPAI obligations effective Aug 2025; high-risk rules staged 2026–2027; enforcement timeline affirmed publicly. [S4]
- D4. Enterprise buyer priorities for AI in HR** — Accuracy, reliability, transparency, and cross-functional governance with IT/legal/compliance. [S6]
- D5. Demonstrated bias risks in LLM-mediated HR tasks** — Evidence of cultural/scoring disparities; necessitates mitigation/monitoring. [S5]
- D6. Capital environment: selective HR-tech funding** — Post-2021 reset; 2024 low base (~\$2B) with cautious recovery. [S7]
- D7. Internal capability (LLM-driven skill-gap analytics)** — Platform scans JDs, performance data, and training catalogs to output gaps and reskilling paths. (Internal description; public validation: Unknown.)

### 3) SWOT

#### A) Strengths (internal)

- Purpose-built skills-gap engine across JDs, performance, and training — Why it matters: Directly addresses structural shortage by aligning demand (roles) with supply (skills) and learning paths — Link: [D1, D2, D7]. Evidence: Unknown.
- Enterprise workflow orientation (HR + L&D + line) — Why it matters: Maps to buyer priorities (skills focus, clear value cases) and supports adoption — Link: [D2, D4, D7]. Evidence: [S6].

Figure D.12: Stage 5 – Grounded in source snippets so claims are evidence based, part 3

- Data network effects potential (richer skills ontology over time) — Why it matters: Improves match quality and recommendation precision with scale — Link: [D2, D7]. Evidence: Unknown.
- B) Weaknesses (internal)**
- Model risk & fairness tooling maturity uncertain — Why it matters: HR buyers scrutinize bias/traceability; gaps slow deals — Link: [D4, D5, D7]. Evidence: [S5][S6].
  - Compliance readiness for EU AI Act unclear — Why it matters: GPAI and (potential) high-risk obligations require documentation, monitoring, and governance — Link: [D3, D4, D7]. Evidence: [S4][S6].
  - Limited brand and reference customers (typical of scale-ups) — Why it matters: In a selective funding and cautious-buyer market, proof points are essential to win enterprise deals — Link: [D6, D4, D7]. Evidence: Unknown.
- C) Opportunities (external)**
- EU-wide digital skills gap and reskilling mandates — Why it matters: Large, policy-reinforced demand for solutions that close skill gaps and accelerate upskilling — Link: [D1, D2]. Evidence: [S1][S2][S3].
  - Buyers prioritizing skills-based talent practices — Why it matters: Budget and executive focus on skills taxonomies, internal mobility, and L&D ROI — Link: [D2, D4]. Evidence: [S2][S6].
  - Compliance-as-a-feature — Why it matters: Proactive alignment with AI Act (documentation, risk management, transparency) can differentiate vs. slower incumbents — Link: [D3, D4]. Evidence: [S4][S6].
  - Consolidation & partnerships in HR-tech — Why it matters: Selective funding climate favors partnerships/reseller routes and potential strategic M&A — Link: [D6]. Evidence: [S7].
- D) Threats (external)**
- Regulatory non-compliance risk — Why it matters: AI Act timelines create delivery risk and potential penalties if obligations unmet — Link: [D3]. Evidence: [S4].
  - Model bias and reputational risk — Why it matters: Documented LLM biases in hiring-adjacent tasks can trigger customer pushback and legal exposure — Link: [D5, D4]. Evidence: [S5][S6].
  - Budget scrutiny and longer enterprise sales cycles — Why it matters: Post-peak HR-tech funding and cautious spending can elongate time-to-close — Link: [D6]. Evidence: [S7].
  - Incumbents embedding similar capabilities — Why it matters: HCM suite vendors can bundle skills analytics, compressing willingness to pay — Link: [D4]. Evidence: [S6].

---

## 4) Strategic Implications

- **Lead with “Responsible, Skills-First AI”** — Productize bias detection, audit trails, and explainability tied to the AI Act’s documentation/monitoring expectations; make this a core value prop, not an add-on (S4, S5, S6). Links: Strengths↔Opportunities/Threats (S1–S6) — [D2, D3, D4, D5, D7].
- **Monetize the reskilling mandate** — Package outcomes (time-to-competency, skill coverage) with CFO-friendly ROI for EU skills targets; co-sell with L&D and business owners (S1–S3, S6). Links: Strengths/Weaknesses↔Opportunities — [D1, D2, D4, D7].
- **De-risk go-to-market in a selective capital climate** — Pursue OEM/ISV and SI partnerships to shorten sales cycles and create reference logos; keep an M&A-ready compliance and telemetry posture (S6, S7). Links: Weaknesses↔Opportunities/Threats — [D4, D6, D7].

Figure D.13: Stage 5 – Grounded in source snippets so claims are evidence based, part 4

## 5) Source List

- [S1] *Commission publishes 2024 State of the Digital Decade Report*, European Commission / Digital Skills & Jobs Platform, 2024 — <https://digital-skills-jobs.europa.eu/en/latest/news/commission-publishes-2024-state-digital-decade-report> ↗
- [S2] *The Future of Jobs Report 2025*, World Economic Forum, 2025 — <https://www.weforum.org/publications/the-future-of-jobs-report-2025/> ↗ and summary article (Jan 8, 2025) — <https://www.weforum.org/stories/2025/01/future-of-jobs-report-2025-jobs-of-the-future-and-the-skills-you-need-to-get-them/> ↗
- [S3] *A new future of work: The race to deploy AI and raise skills in Europe and beyond*, McKinsey Global Institute, 2024 — <https://www.mckinsey.com/mgi/our-research/a-new-future-of-work-the-race-to-deploy-ai-and-raise-skills-in-europe-and-beyond> ↗ (PDF variant available)
- [S4] *AI Act | Shaping Europe's digital future* (application timeline) & *Official Journal* publication info, European Commission, 2024–2025 — <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai> ↗ and consolidated act text — <https://artificialintelligenceact.eu/the-act/> ↗
- [S5] *Invisible Filters: Cultural Bias in Hiring Evaluations Using Large Language Models*, Rao, P. S. B. et al., arXiv, 2025 — <https://arxiv.org/abs/2508.16673> ↗ (HTML/PDF)
- [S6] *AI in HR: How AI Is Transforming the Future of HR*, Gartner (public topic page), 2025 — <https://www.gartner.com/en/human-resources/topics/artificial-intelligence-in-hr> ↗
- [S7] *Funding To HR Software Startups Rises As M&A Activity Picks Up*, Crunchbase News, Sep 19, 2025 — <https://news.crunchbase.com/venture/ai-hr-software-startup-funding-ma/> ↗

### Notes:

- Internal capability assertions about the specific scale-up are marked “Unknown” when no public evidence is available.
- EU AI Act enforcement cadence reinforced by Commission materials; GPAI obligations effective Aug 2, 2025, and high-risk rules phased through 2026–2027.

[S4]

   ...  Bronnen

## **D.6 Stage 6 – Targeted refinement & finalization**

The model is asked to self-assess the draft output against a simple rubric (structure adherence, completeness, clarity of implications, explicit assumptions) and revise once. This produces the final illustrative artifact with improved clarity and transparency.

Figure D.14: Stage 6 – Quality checking the output of the LLM and refining it, part 1

You are reviewing the draft SWOT analysis produced above.  
Do not regenerate the SWOT immediately. Instead, follow these steps:

**Step 1 – Critique:**  
Critique the draft against the following rubric in 3–4 bullets:

- Structure adherence: Did it follow the required format (Key Drivers, SWOT, Implications, Source list)?
- Completeness: Are all quadrants distinct and populated with non-overlapping, specific items?
- Clarity: Are the SWOT bullets crisp and actionable, avoiding vague phrases? Do these bullets really provide strategic value to the company and consultant?
- Assumptions & uncertainties: Are data gaps or limitations made explicit?

**Step 2 – Prompt improvement:**  
Based on the critique, write a revised prompt that would likely produce a stronger output.  
The new prompt should keep the same role, task, and context as above, but adjust the instructions to address weaknesses you identified.

**Step 3 – Display prompt:**  
Display the refined prompt for transparency.

**Step 4 – Execute:**  
Now execute your improved prompt once, producing a revised SWOT analysis (including Key Drivers, SWOT quadrants, Strategic Implications, and Source List).



Nagedacht

Here's a tight QC pass and an improved run.

### Step 1 – Critique

- **Structure adherence:** Generally followed the requested sections; however, the rubric asked that *each SWOT bullet* include [D#] links and, when evidence-based, an inline [S#] tag. The draft sometimes put "Evidence:" after the bullet rather than embedding [S#] inside the bullet body, and the "Key Sources" list exceeded the "5–7 sources" guidance.
- **Completeness (non-overlap & specificity):** Quadrants are mostly distinct, but some items are generic (e.g., "Enterprise workflow orientation") and a few threats/opportunities blur (funding climate vs. sales cycles). Competitive dynamics are mentioned without concrete, cited examples (e.g., named suite vendors' moves).
- **Clarity & actionability:** Many bullets are crisp, but several rely on vague benefits ("data network effects potential") without KPIs or buyer-relevant proof points. "Compliance-as-a-feature" is sound but would benefit from concrete AI Act obligations (e.g., GPAI documentation, Annex III employment use cases) and how to productize them.
- **Assumptions & uncertainties:** "Unknown" is used to flag internal claims—which is good—but the draft could more explicitly quantify key drivers (e.g., % skill change, funding levels) and anchor competitive claims with public events (acquisitions/launches).

### Step 2 – Prompt improvement

Revised prompt (for the model generating the SWOT):

You are a strategy consultant preparing a client-ready note for senior executives.

Figure D.15: Stage 6 – Quality checking the output of the LLM and refining it, part 2

**Role, task, context:** Same as before (EU HR-tech scale-up building a GenAI workforce-planning service that scans JDs, internal performance data, and training catalogs to output skill-gap analyses and reskilling strategies; target: digital/engineering shortages).

**Do this exactly:**

1. **“Key Sources (5–7 only):** Select high-quality, recent public sources (e.g., EU Digital Decade, WEF Future of Jobs 2025, McKinsey MGI 2024, EU AI Act timeline, OECD skills training, Crunchbase News on HR-tech funding, a recent peer-reviewed/preprint on LLM bias in hiring). Write 1–2 line summaries with a concrete stat or finding per source. Tag each as [S#].”
2. **“Key Drivers (5–8):** Mix external and internal drivers. Each driver must include a number/date or named obligation where applicable (e.g., “39% skill instability by 2030,” “GPAI obligations at +12 months post-EIF”). Tag each with its source [S#].”
3. **“SWOT:** For each bullet, use the schema  
**Claim — Why it matters — Link: [D#,...] [S#?]”**
  - “Include at least one [D#] per bullet.”
  - “Include [S#] for any evidence-based statement inside the bullet.”
  - “Avoid overlap; make bullets specific (KPIs, buyer criteria, obligations).”
  - “Include one competitive bullet citing a concrete market event (e.g., acquisition).”
  - “Mark internal, non-public assertions as “Unknown.””
4. **“Strategic Implications (2–3 bullets):** Explicitly connect S/W ↔ O/T with clear actions (e.g., “ship model cards and bias audits aligned to AI Act GPAI docs; target buyer X; KPI Y”). Include [S#] where relevant.”
5. **“Source List: Map [S#] to full reference (title, author/organization, year, URL).”**

**Style constraints:** Professional consulting tone; American English; crisp sentences; quantify where possible; no more than 7 sources total.

### Step 3 – Display prompt

(Exact text you can reuse)

**Prompt to execute:**

[Use the “Revised prompt” from Step 2 above verbatim.]

### Step 4 – Execute (revised SWOT)

#### 1) Key Sources

- [S1] European Commission, *State of the Digital Decade 2024*: Only **55.6%** of EU citizens have at least basic digital skills; ICT specialists on track for **~12M vs. 20M** target by 2030—evidence of a persistent EU skills gap. Digital Skills and ...
- [S2] World Economic Forum, *Future of Jobs 2025*: Employers expect **39%** of workers’ skills to be transformed/outdated by 2025–2030; AI/big data among fastest-growing skills; **59%** of workforce needs training by 2030. World Economic ...
- [S3] McKinsey Global Institute (2024), *A new future of work*: Demand rising for **STEM** and high-skill roles; up to **27–30%** of hours automated by 2030; European firms plan to emphasize **retraining**. McKinsey & Com...
- [S4] Future of Life Institute, *EU AI Act — High-level summary & timelines*: **+12 months** after entry into force for GPAI obligations; **+24 months** for **Annex III high-risk** systems (employment, worker management listed). artificialintelligen...
- [S5] OECD (2025), *Bridging the AI skills gap*: Only **0.3–5.5%** of analyzed training courses contain AI content; countries ramping up AI literacy and professional training but supply still lags demand. OECD
- [S6] Crunchbase News (Sept 19, 2025): HR software startups raised **\$1.9B YTD 2025 vs. \$2.0B in 2024**; **Workday** announced acquisitions of **Sana (\$1.1B)** and **Paradox**—evidence of consolidation and incumbents bundling AI. Crunchbase News
- [S7] Anzenberg et al. (2025), *Evaluating the Promise and Pitfalls of LLMs in Hiring*: Benchmarks show **LLMs can exhibit demographic disparities** unless mitigated; calls for domain-specific models and bias audits. arXiv

Figure D.16: Stage 6 – Quality checking the output of the LLM and refining it, part 3

## 2) Key Drivers

- D1. EU digital skills gap remains large** (55.6% basic skills; ICT shortfall vs. 2030 targets). [S1] [Digital Skills and ...](#)
- D2. Skill instability accelerates reskilling demand** (39% of skills change by 2030; 59% need training). [S2] [World Economic ...](#)
- D3. AI Act compliance clock** (+12 months GPAI; +24 months Annex III/high-risk including employment/worker management). [S4] [artificialintelligen...](#)
- D4. Enterprises plan retraining, not just hiring** (rising STEM demand; 27–30% hours automated by 2030). [S3] [McKinsey & Com...](#)
- D5. Fairness & auditability are table-stakes** (evidence of bias in hiring-adjacent LLM use without safeguards). [S7] [arXiv](#)
- D6. Funding selective; consolidation rising** (\$1.9B YTD 2025; Workday acquiring AI HR tech). [S6] [Crunchbase News](#)
- D7. Training supply constraints** (only 0.3–5.5% of courses include AI; need to scale AI literacy). [S5] [OECD](#)

## 3) SWOT

### A) Strengths (internal)

- **Purpose-built skills-gap & reskilling recommender across JDs, performance, and training catalogs** — Why it matters: Directly targets EU skill shortages and 39% skill instability, enabling precise learning paths for digital/engineering roles — Link: [D1, D2] **Evidence:** Unknown.
- **Tight focus on digital & engineering roles** — Why it matters: Aligns to rising demand for STEM/high-skill roles and employer plans to retrain into these areas — Link: [D2, D4] [S3]. [McKinsey & Com...](#)
- **Potential data flywheel (skills ontology improves as outcomes/usage grow)** — Why it matters: Better gap detection and recommendations over time, improving adoption and ROI — Link: [D2, D7] **Evidence:** Unknown.

### B) Weaknesses (internal)

- **Fairness, explainability, and audit tooling maturity unclear** — Why it matters: HR buyers will expect bias testing, audit trails, and documentation; gaps risk stalled pilots — Link: [D3, D5] [S4][S7]. [artificialintelligen... +1](#)
- **Compliance readiness for AI Act timelines uncertain** — Why it matters: GPAI (+12 months) and Annex III (employment/worker management, +24 months) impose documentation, risk management, and oversight — Link: [D3] [S4]. [artificialintelligen...](#)
- **Outcomes depend on client data and training catalog quality** — Why it matters: If enterprise training supply is thin, realized reskilling throughput may lag expectations — Link: [D7] [S5]. [OECD](#)

### C) Opportunities (external)

- **Policy-reinforced EU reskilling mandate** — Why it matters: Large, urgent demand to close digital skills gaps across member states; clear buyer funding and executive attention — Link: [D1, D2] [S1][S2]. [Digital Skills and ... +1](#)
- **Enterprise shift toward retraining for AI adoption** — Why it matters: Budgeted programs and KPIs around internal mobility and time-to-competency favor solutions that map gaps to courses — Link: [D4] [S3]. [McKinsey & Com...](#)
- **Compliance-as-a-feature differentiation** — Why it matters: Packaging GPAI documentation, model cards, risk management, and human-oversight enablers can win legal/IT stakeholders — Link: [D3] [S4]. [artificialintelligen...](#)
- **Partner & OEM routes amid consolidation** — Why it matters: Incumbents are acquiring/bundling AI capabilities (e.g., Workday–Sana, Workday–Paradox), opening BD/OEM and exit paths — Link: [D6] [S6]. [Crunchbase News](#)

### D) Threats (external)

- **Regulatory non-compliance risk on EU AI Act timelines** — Why it matters: Missed GPAI/Annex III obligations can block deployments and trigger penalties — Link: [D3] [S4]. [artificialintelligen...](#)

Figure D.17: Stage 6 – Quality checking the output of the LLM and refining it, part 4

- **Bias & reputational exposure in hiring-adjacent analytics** — Why it matters: Documented disparities in LLM outputs can erode trust and invite scrutiny unless mitigated — Link: [D5] [S7]. [arXiv](#)
- **Incumbent HCM vendors bundling skills/AI** — Why it matters: Bundles compress willingness-to-pay and increase table-stakes features — Link: [D6] [S6]. [Crunchbase News](#)
- **Training supply bottlenecks slow impact** — Why it matters: If client catalogs lack AI/digital content, recommended pathways won't convert into real skill uplift — Link: [D7] [S5]. [OECD](#)

#### 4) Strategic Implications

- **Productize “Responsible Skills AI”** — Ship model cards, GPAI docs, audit logs, and bias-testing reports mapped to AI Act requirements (GPAI + Annex III employment/worker-management) and make them buyer-visible (Legal/IT). Prioritize **explainable gap reasons** and **human-in-the-loop controls**. *KPIs*: win rate in regulated industries; % deals with compliance sign-off at first security review. [S4][S7] — Links: Strengths (skills engine) ↔ Opportunities (compliance) & Threats (regulatory/bias). [artificialintelligen... +1](#)
- **Monetize reskilling outcomes, not seats** — Package offers around **time-to-competency**, % **skill coverage closed**, and **internal fill rate for STEM roles**, tying to enterprise retraining strategies and WEF skill instability metrics. Co-sell with L&D + business unit leaders. [S2][S3] — Links: Strengths (skills engine) & Weakness (data dependency) ↔ Opportunities (policy-reinforced reskilling). [World Economic ... +1](#)
- **Go-to-market via partners to counter bundling** — Pursue **OEM/ISV integrations** with LMS/LXP and SI partners; position as the **skills-intelligence layer** inside incumbent HCM stacks; build case studies in EU manufacturing/engineering. Track **pipeline sourced via partners** and **attach rates**. [S6] — Links: Weakness (brand/footprint) ↔ Threats (incumbent bundling) & Opportunities (consolidation routes). [Crunchbase News](#)

#### 5) Source List

- [S1] *Commission publishes 2024 State of the Digital Decade Report*, European Commission / Digital Skills & Jobs Platform, 2024 — <https://digital-skills-jobs.europa.eu/en/latest/news/commission-publishes-2024-state-digital-decade-report> [Digital Skills and ...](#)
- [S2] *The Future of Jobs Report 2025 – Key Findings*, World Economic Forum, 2025 — <https://www.weforum.org/publications/the-future-of-jobs-report-2025/digest/> [World Economic ...](#)
- [S3] *A new future of work: The race to deploy AI and raise skills in Europe and beyond*, McKinsey Global Institute, 2024 — <https://www.mckinsey.com/mgi/our-research/a-new-future-of-work-the-race-to-deploy-ai-and-raise-skills-in-europe-and-beyond> [McKinsey & Com...](#)
- [S4] *EU Artificial Intelligence Act — High-level summary & timelines*, Future of Life Institute, 2024–2025 — <https://artificialintelligenceact.eu/high-level-summary/> [artificialintelligen...](#)
- [S5] *Bridging the AI skills gap: Is training keeping up?*, OECD, 2025 — [https://www.oecd.org/content/dam/oecd/en/publications/reports/2025/04/bridging-the-ai-skills-gap\\_b43c7c4a/66d0702e-en.pdf](https://www.oecd.org/content/dam/oecd/en/publications/reports/2025/04/bridging-the-ai-skills-gap_b43c7c4a/66d0702e-en.pdf) [OECD](#)
- [S6] *Funding To HR Software Startups Rises As M&A Activity Heats Up*, Crunchbase News (Mary Ann Azevedo), Sept 19, 2025 — <https://news.crunchbase.com/venture/ai-hr-software-startup-funding-ma/> [Crunchbase News](#)
- [S7] Anzenberg, Samajpati, Chandrasekar & Kacholia (2025), *Evaluating the Promise and Pitfalls of LLMs in Hiring Decisions*, arXiv — <https://arxiv.org/abs/2507.02087> [arXiv](#)

Notes: Internal capabilities are labeled “Unknown” where no public evidence is provided.

# Chapter E Figures

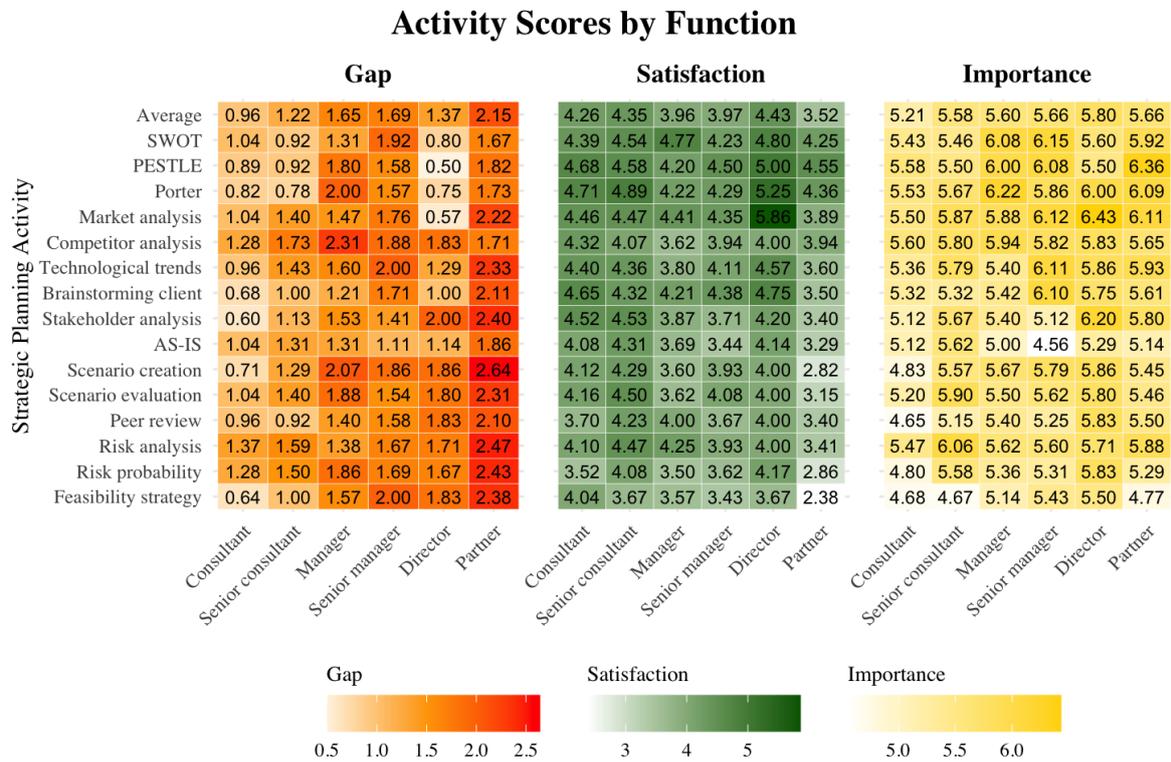


Figure E.1: Heatmap of gap, satisfaction, and importance scores for each strategic planning activity, broken down by functional role. Values represent mean scores per function. Darker shades indicate more extreme values within each scale.

## Activity Scores by Experience

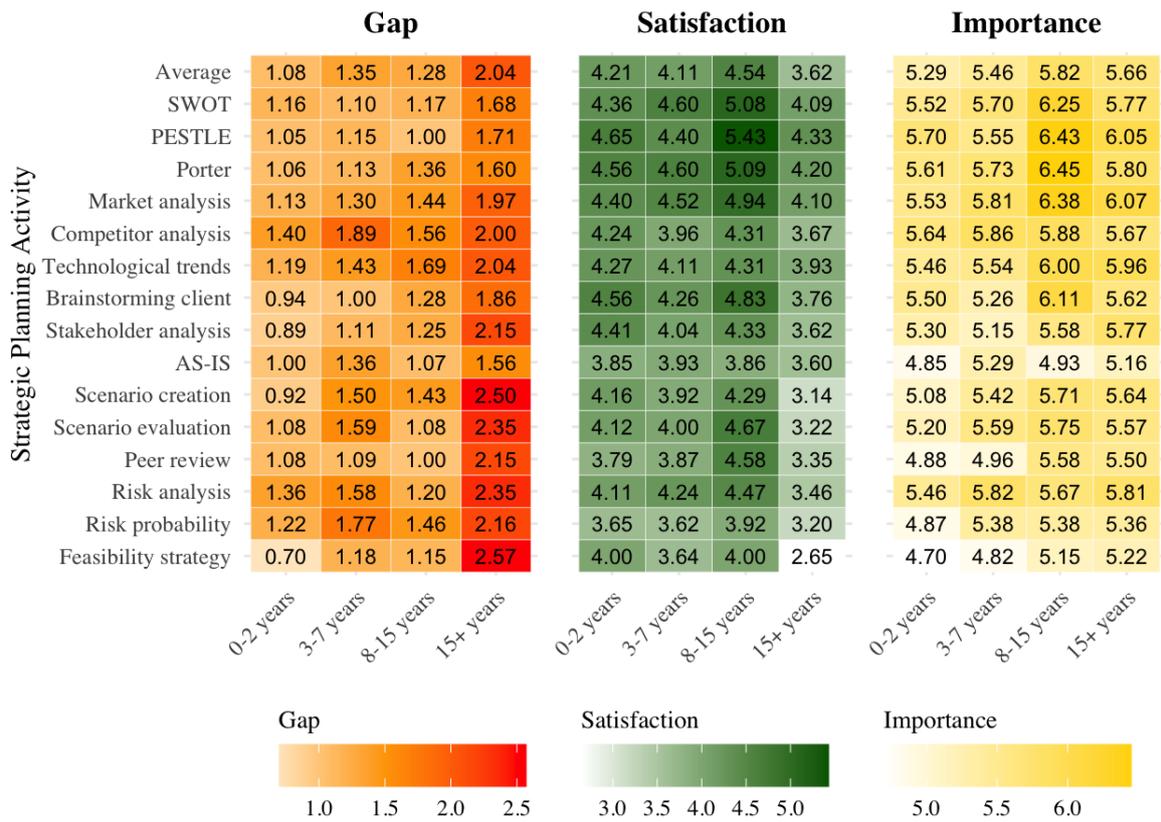


Figure E.2: Heatmap of gap, satisfaction, and importance scores for each strategic planning activity, broken down by experience group. Values represent mean scores per experience group. Darker shades indicate more extreme values within each scale.

## Activity Scores by Department

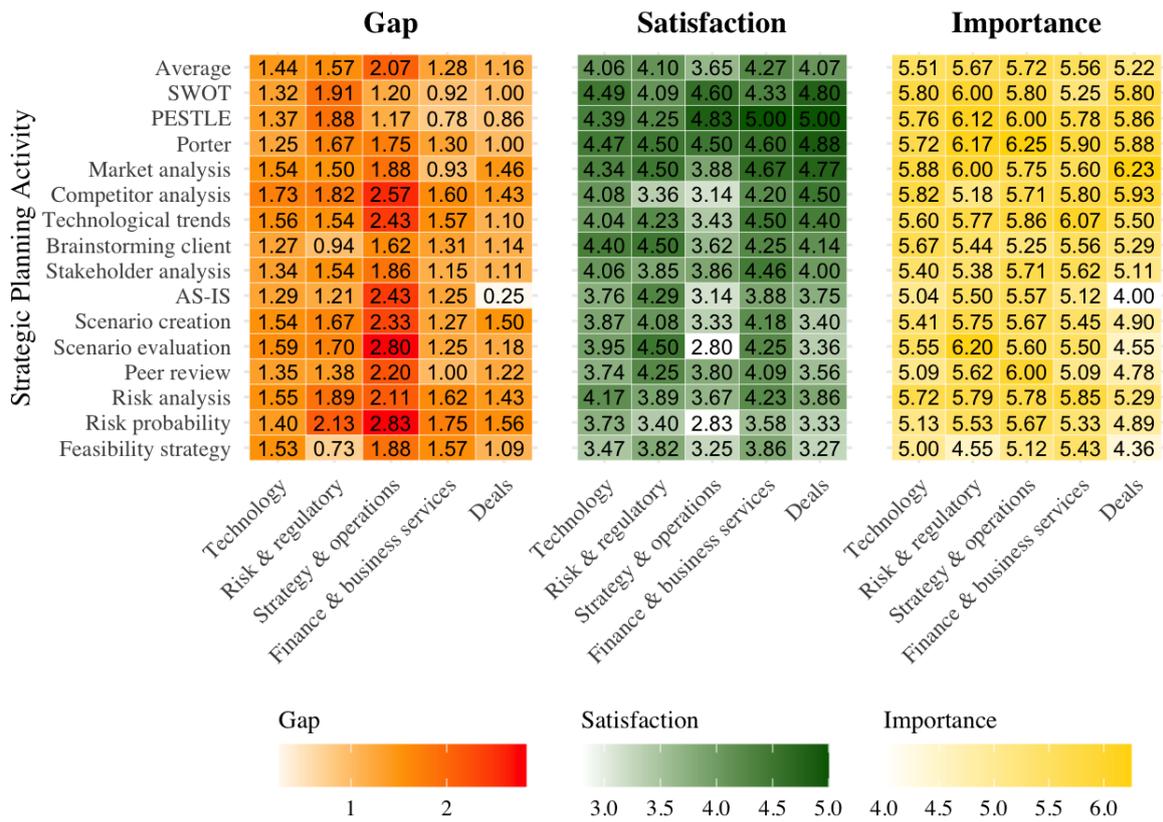


Figure E.3: Heatmap of gap, satisfaction, and importance scores for each strategic planning activity, broken down by department. Values represent mean scores per department. Darker shades indicate more extreme values within each scale.