Vaasan yliopisto
UNIVERSITY OF VAASA

Konsta Ahola

The impact of generative artificial intelligence on management consulting work processes and productivity

School of Technology and Innovations
Master of Science in Economics and Business Administration
Master’s thesis in Industrial management

Vaasa 2026

VAASAN YLIOPISTO
School of Technology and Innovations
Tekijä: Konsta Ahola
Tutkielman nimi: The impact of generative artificial intelligence on management consulting work processes and productivity
Tutkinto: Kauppatieteiden maisteri
Koulutusohjelma: Tuotantotalous
Työn ohjaaja: Jyri Naarmala

Valmistumisvuosi: 2026 **Sivumäärä:** 72

TIIVISTELMÄ:
Liikkeenjohdon konsultointiyritykset tuottavat arvoa analyysin, tulkinnan ja asiakaskohtaisen ongelmanratkaisun kautta. Generatiivisesta tekoälystä on tullut osa konsultoinnin työprosessia, sillä se kykenee auttamaan useissa työn keskeisissä tehtävissä. Tässä tutkielmassa tarkastellaan, kuinka generatiivinen tekoäly vaikuttaa liikkeenjohdon konsultoinnin työprosesseihin ja millä edellytyksillä tehtävätason tehokkuushyödyt voivat muuttua työn tuottavuuden kasvuksi.

Tutkimus perustuu integratiiviseen kirjallisuuskatsaukseen ja olemassa olevan empiirisen aineiston temaattiseen analyysiin. Aineisto koostuu kokeellisista tutkimuksista, generatiivisen tekoälyn käyttöä koskevista kyselytutkimuksista sekä toimialaraporteista, ja se nojaa erityisesti ihmisen ja tekoälyn yhteistyötä sekä asiantuntijatyön tuottavuutta käsittelevään kirjallisuuteen.

Tutkimuksen perusteella selkeimmät tekoälyn hyödyt liittyvät varhaisen vaiheen analyyttisiin tehtäviin ja kirjoittamiseen, kun taas päätöksenteossa ja ongelmanratkaisussa vaikutukset ovat epäselvemmät. Tätä selittää niin sanottu epätasainen teknologinen rajapinta: sama työkalu voi toimia hyvin joissakin tehtävissä ja huonosti toisissa. Tekoälyn tuottama teksti voi myös vaikuttaa laadukkaalta, vaikka sen taustalla oleva analyysi olisi puutteellinen. Siksi tuotosten huolellinen tarkistaminen on kasvavissa määrin tärkeä osa konsultointityötä. Työn nopeutuminen tehtävätasolla ei automaattisesti tarkoita, että koko projekti muuttuu tuottavammaksi. Säästetty aika täytyy käyttää laadukkaampaan analyysiin ja asiakkaan tavoitteiden saavuttamiseen, jotta työstä tulee kokonaisuudessaan tuottavampaa.

AVAINSANAT: generatiivinen tekoäly, ihmisen ja tekoälyn yhteistyö, konsultoinnin tuottavuus, liikkeenjohdon konsultointi, tietotyö

UNIVERSITY OF VAASA
School of Technology and Innovations
Author: Konsta Ahola

Title of the thesis:	The impact of generative artificial intelligence on management consulting work processes and productivity
Degree:	Master of Science in Economics and Business Administration
Degree Programme:	Industrial management
Supervisor:	Jyri Naarmala
Year:	2026 **Pages:** 72

ABSTRACT:

Generative artificial intelligence has become part of knowledge-intensive professional services at a point when its capabilities seem to align with many key consulting tasks. Management consulting firms add value through analysis, interpretation, and problem-solving that is specific to each client; however, it is still unclear whether GenAI truly increases consulting productivity. This thesis focuses on how generative AI affects management consulting work processes and under what conditions task-level efficiency gains can translate into realised consulting productivity.

The research method in this study is integrative literature review and thematic analysis of secondary empirical evidence, which include experimental studies of consulting-type tasks, surveys on generative artificial intelligence adoption, and industry reports.

The findings suggest that GenAI delivers the clearest benefits in early-stage analysis and writing. Decision-making and problem-solving prove more challenging, as the same tool that performs well in some tasks can be unreliable in others. GenAI outputs can also appear polished despite flawed underlying reasoning, which makes verification an important part of responsible use.

Higher consulting productivity does not follow automatically from task-level efficiency gains. Time savings need to be directed towards better analysis and sounder judgement to produce any real benefit. GenAI can support productivity in management consulting, but only when integrated deliberately into firm workflows, used within clear guidelines, and accompanied by consultant accountability for the final output.

KEYWORDS: generative artificial intelligence, management consulting, knowledge work, consulting productivity, human–AI collaboration

Contents

Figures

Tables

Abbreviations

GenAI	**Generative Artificial Intelligence**
AI	**Artificial Intelligence**
LLM	**Large Language Model**
EDPS	**European Data Protection Supervisor**
OECD	**Organisation for Economic Co-operation and Development**
BCG	**Boston Consulting Group**

Artificial intelligence tools were used during the thesis process for research design ideation, thesis structure planning, language revision and for identifying how to progress in some part of the writing process. OpenAI's ChatGPT and Anthropic's Claude were used to support the writing process.

In this thesis, Artificial Intelligence was not used to generate any research data or as a replacement for academic sources. Sources used in this thesis were verified from the original publications. The author is responsible for all content and interpretations presented in this study.

1 Introduction

Management consulting firms are knowledge-intensive professional service firms whose work centres on analysis, writing, and structured communication. Generative AI technologies have entered this environment at a time when their capabilities fit well with many core consulting activities. Whether these capabilities translate into realised productivity remains an open question, which this thesis addresses. This chapter introduces the background of the study, the research gap, the research question, and the scope of the thesis, followed by an overview of the thesis structure.

1.1 Background of the study

The connection between generative AI and knowledge work lies partly in speed. Generative AI produces and organises language-based content faster than earlier digital technologies. At the workplace level, this can be seen in everyday activities like writing, document summarisation, communication preparation, idea generation, and analysis. While these activities may seem routine, they account for a significant proportion of professional service work. They also represent how knowledge workers transform information into something usable by others. This thesis focuses specifically on generative AI rather than artificial intelligence in general. Artificial intelligence is a broad category of technologies that can be defined as a form of automation based on pattern recognition from data inputs. The term 'generative artificial intelligence' refers to algorithms capable of creating new content—such as text, images, code, or structured responses—based on patterns learned from data (Feuerriegel et al., 2024). Text- and knowledge-related applications are the central theme of this research.

Generative AI is particularly important for the management consulting industry. Consulting work is a knowledge-intensive professional service that creates value by applying expertise to solve a client's specific organisational challenges. Consultancies produce deliverables based on analysis, synthesis, reporting, communication, decision support, and problem-solving. In these tasks, information must be processed, analysed, and used as a basis for making judgements within specific consulting projects.

The existing literature describes management consulting as a specific professional activity requiring both the application of expert advice and client collaboration (Mosonyi et al., 2020). Consulting is further characterised by high knowledge intensity and a reliance on expert labour (von Nordenflycht, 2010). At the same time, the industry uses codified resources such as methods, frameworks, templates, and past case materials. GenAI can support the work done by consulting firms, as it can operate with this kind of codified, language-based knowledge.

The complexity of consulting means that productivity is not easily measurable. Unlike some other industries, consulting cannot rely simply on measures of output per hour or output cost. A longer recommendation or a larger number of slides is not necessarily better than a concise one. Speed in producing a report is not a sufficient measure of good consulting work. This is why consulting productivity must also include considerations of accuracy, relevance, and outcomes in client organisations (Palvalin, 2019; Óskarsdóttir et al., 2022).

These characteristics point toward the principal issue of this study. GenAI might facilitate certain consulting tasks by helping to produce drafts, summarise text, compare options, or prepare communication. However, the productivity of consulting cannot be measured only by speed in completing those tasks. Rather, it depends on accurate analysis, applicable recommendations, defensible conclusions, and actions taken by the client regarding the consulting project. This means that GenAI is likely to have conditional impacts on consulting work and consulting tasks. The idea of the jagged technological frontier is applicable here. According to Dell'Acqua et al. (2023), AI technologies can boost efficiency in some tasks while weakening performance in others, often in ways that are difficult to predict. In consulting, this is especially relevant because the process includes a wide range of activities, from drafting texts to making judgements about clients' business cases. A consultant might use GenAI to draft an initial structure for a slide deck or to generate summary texts. However, GenAI is less likely to support diagnostic and judgement-based tasks related to analysing the client's situation. Experimental studies show that generative AI can have visible task-level impacts, especially on speed. However, a fast draft can still create rework if it requires heavy correction, and a generated recommendation remains a risk if it is not grounded in solid evidence.

1.2 Research gap and research question

The current academic literature suggests that the use of generative AI might lead to enhanced performance in selected knowledge-work tasks. For example, Dell'Acqua et al. (2023) conduct a field experiment and find direct evidence related to consultants' productivity improvement by performing consulting tasks. Noy and Zhang (2023) find positive impacts on productivity for professional writing tasks, an important result given that consultants spend significant amounts of time writing. Finally, Brynjolfsson et al. (2023) provide evidence from the customer support sector. These studies suggest what outcomes AI adoption might produce in task settings.

Much of this evidence is limited to specific tasks within controlled or constrained environments. While a writing task, a product-innovation task, or a customer-support interaction can demonstrate significant effects, none fully reflect the complexities of a management consulting project. Consulting involves project-based, client-specific, and iterative processes. A team often starts with unclear information, develops hypotheses, tests them with client input, revises the scope, and ultimately delivers a recommendation. Therefore, task-level findings offer relevant context but do not alone explain how GenAI impacts an entire consulting workflow.

On another side, the existing literature demonstrates broader impacts. According to Bick et al. (2025), the rate of AI adoption is high in general, but the amount of work hours affected remains relatively small. McKinsey & Company (2023), Accenture (2024), and Deloitte (2024, 2025) offer insights about AI productivity potential, organisational adoption, and scaling issues. The sources show how organisations see the impact of AI, but should be analysed critically. The consulting organisations discuss the potential, not actual productivity change.

The gap lies between these two types of evidence. Experimental evidence shows speed improvements or quality increases within tasks. Industry reports describe large organisational potential. The issue that remains unclear is how gains at a task-level translate into consulting work processes and consulting productivity. It is possible for a task to become quicker and yet leave a project completion speed unchanged; in other words, one can draft a document quickly, but consultants still have to check it, adjust it, and implement it.

This consideration is especially relevant given that management consulting combines information production with the provision of credible advice to clients. This is why there are intermediate factors between efficiency gains at the task level and consulting productivity. These include review, interpretation, workflow design, adoption level, and governance, among others. Although the literature mentions these factors in passing, very few studies explicitly connect them to the realities of management consulting projects.

The research question of this thesis is:
How does generative AI affect management consulting work processes, and under what conditions can task-level efficiency gains translate into realised consulting productivity?

The question has two components. The first considers how GenAI influences activities like analysis, initial structuring, professional writing, reporting, decision support, and problem-solving. The second question asks when improvements in speed or output quality at the local level translate into meaningful effects at the project or organisational scale. The thesis views GenAI's effect as conditional, depending on the task type, review quality, and how consulting teams incorporate the tool into their workflows.

1.3 Objectives and scope of the study

The thesis has four main objectives. The first objective is to explain how GenAI influences consulting processes through information management, summarisation, report writing, presentation slide creation, and decision-making. The focus is not on all possible applications of AI in consulting firms, but on the processes used to deliver consulting projects. The second objective involves assessing productivity effects at different levels. Productivity at the task level refers to how efficiently a specific task, such as summarising material, can be completed. Productivity at the project level relates to whether the entire consulting assignment improves in terms of speed, cost-effectiveness, risk reduction, or value creation for the client. Productivity at the firm level is broader and relates, for example, to the scalability of GenAI use or knowledge creation. These levels remain separate in the thesis, since findings at one level do not necessarily imply similar findings at another.

The third objective is to explain why time savings do not necessarily equate to consulting productivity. Productivity, in this case, means more than being fast. It can mean producing high-quality results, producing something relevant to the client’s needs, reducing risks, shortening lead times, and ensuring that the work supports more effective decision-making. GenAI may result in faster writing or synthesis, but its effect still depends on what follows after the first output is generated.

The fourth objective is to identify the key determinants of GenAI’s value in consulting. These include validation processes, contextualisation, workflow design, level of adoption, governance challenges, and consultant judgement. Related risks include overdependence, hallucinations, inadequate data security, and the impact of such technology on the learning process of junior consultants.

The scope of this thesis is restricted to management consulting and related knowledge-work settings. There is limited direct empirical evidence on GenAI in management consulting; therefore, relevant findings from adjacent contexts are also included in this thesis. These sources can be considered analogue evidence related to consulting activities rather than direct evidence of consulting outcomes.

This research does not cover technical issues related to artificial intelligence, including model design, learning algorithms, training data, or software engineering. It also does not involve primary data collection. The study applies an integrative literature review approach based on thematic analysis of secondary evidence and uses various types of literature, including conceptual, experimental and quasi-experimental, adoption, survey, governance, and selected consulting industry literature.

1.4 Structure of the thesis

The thesis is organised into five chapters. The first chapter of this thesis introduces the area of focus, identifies the research gap, formulates the research question, and presents the objectives and scope of the study. The second chapter sets out the theoretical basis by defining key concepts such as artificial intelligence, generative artificial intelligence, knowledge work, and productivity in knowledge-intensive work. It also includes human–AI interaction, productivity in AI-enabled knowledge work, and the risks involved in such interaction. Chapter 3 outlines the methodology, which uses an integrative literature review approach and a thematic synthesis of secondary empirical evidence. It describes how the material was collected, selected, and analysed; it presents the three-part evidence structure used to compare experimental studies, adoption evidence, and industry reports; and it identifies the limitations of the research. Chapter 4 provides an empirical evidence synthesis of how GenAI impacts analysis, structuring, professional writing, reporting, decision-making, and problem-solving tasks. It explores factors that either promote or hinder productivity improvements, comparing direct consulting evidence with consulting-relevant analogue evidence and considering industry reports as contextual rather than causal. In Chapter 5, the findings are interpreted in relation to theory, with a focus on how improvements in task-level efficiency can result in productivity gains for consulting under specific conditions. The chapter also covers managerial implications, limitations, directions for future research, and concludes the study.

2 Literature Review

Before analysing the empirical evidence, the thesis first needs a theoretical basis. This literature review part defines the main concepts and explains why management consulting is a useful setting for studying GenAI. The focus is not on reviewing all AI or consulting research. Instead, the chapter brings together the concepts that are needed for the research question. It first narrows the discussion from artificial intelligence to generative AI. Then it explains consulting as knowledge-intensive professional service work. After that, the chapter discusses human–AI collaboration, productivity in knowledge work, and the main risks linked to GenAI use. The final section brings these ideas together and further discusses othe research gap.

2.1 Key concepts and scope

Generative AI is a narrower concept than artificial intelligence in general. In this thesis, AI is used as the wider term. The actual focus is on generative AI. Feuerriegel et al. (2024, p. 111) define generative AI as computational techniques that use training data to produce seemingly new and meaningful content, such as text, images, or audio. This focus aligns with the topic of the thesis. Consulting work is mainly based on language and information. It involves generating, transforming, interpreting, and communicating knowledge. It is not mainly about automating physical tasks.

AI is also treated in a limited way in this thesis. The focus is not on AI replacing professional expertise. Instead, the thesis looks at AI tools that are used as part of organisational work. Feuerriegel et al. (2024, p. 116) describe application-level generative AI as systems that are connected to human tasks and organisational needs. These systems are used to solve specific business problems and to support people in particular tasks.

Knowledge work is one of the key concepts in this thesis. However, it should not be treated as equivalent to office work or computer-based work. Pyöriä (2005, pp. 120–124) discusses the non-routine character of tasks, their symbolic content, and the importance of judgement and discretion as central features of knowledge work. Knowledge work is also not a uniform category. Pyöriä (2005, p. 124) notes that knowledge workers vary considerably, which means different knowledge-intensive tasks may be affected by technologies like GenAI to different degrees.

Productivity is difficult to define in knowledge-intensive work. In standardised settings, output-input ratios offer a useful way to measure performance. Professional work is different. Outputs are less tangible, more client-specific, and harder to compare across cases. As stated before, faster does not necessarily mean better. This becomes even more relevant when clients try to evaluate expert work. Von Nordenflycht (2010, pp. 158–161) points out that professional services involve an asymmetry of expertise. Clients may not have the background needed to judge the quality of what they receive.

These concepts define a narrow scope for this thesis. The research focuses on generative AI as a technology used in knowledge-intensive professional services and its role in information processing, analysis, reporting, communication, and decision support. Productivity is understood not only in terms of efficiency but also in terms of value creation. The next section examines management consulting in its analytical context.

2.2 Management consulting as knowledge-intensive work

Management consulting involves various activities, so a clear definition is needed for this research topic. Mosonyi et al. (2020) use Greiner and Metzger's definition. Under this definition, management consultants do more than give general advice to clients. They help clients identify management issues, analyse them, generate possible solutions, and sometimes support the implementation of these solutions. This definition is useful for this thesis because GenAI enters consulting work through these stages of the service process.

Based on the above, consulting can be understood as knowledge-intensive work. However, it differs from many other types of knowledge-intensive work because its main input is expertise, not machines or physical capital (von Nordenflycht, 2010, p. 155). Its main raw material is information, and its outputs include analyses, interpretations, and recommendations. To describe this more clearly, it is useful to combine two bodies of literature: professional service firms (PSFs) and knowledge-intensive business services (KIBS).

Consulting depends heavily on expertise, which makes the professional service firm literature relevant here. According to von Nordenflycht (2010, p. 155), professional service firms are characterised by knowledge intensity, low capital intensity, and a professionalised labour force. These features fit consulting well, since consulting work relies strongly on education and knowledge. Expertise also helps create trust between consultants and clients. However, consulting is not a classical profession in the same way as law. Von Nordenflycht (2010, p. 164) states that consulting is less professionalised because there is no professional organisation that regulates entry into the consulting field or defines how consulting firms should be organised.

Knowledge-intensive business services (KIBS) add the service side of consulting. According to Schlee et al. (2024, p. 762), KIBS can be defined as expert services marked by a high degree of customisation and low standardisation. This definition fits consulting well. Even if consultants use frameworks, templates, or other tools, they still have to customise them for the client’s problem.

Client co-production is also important in KIBS research. Den Hertog (2010, pp. 205–206) refers to knowledge-intensive business services as co-producers of innovation, where provider and client work closely together to solve problems. He also names three possible roles for KIBS: facilitator, carrier, and source of innovation (den Hertog, 2010, pp. 206–207). Bettencourt et al. (2002, p. 101) make a similar argument. In complex knowledge-intensive services, clients also take part in solving the problem. They are not only passive receivers of the service. In consulting, the consultant cannot simply take the problem away and return with a finished answer. The work requires access to the client's information, people, and interpretation of the situation.

Werr and Stjernberg (2003, p. 881) offer a practical view of how knowledge is organised inside consulting firms. They argue that consulting firms rely on methods and tools, case knowledge, and individual consultant experience. Methods and tools give structure to the work. Case knowledge keeps earlier project experience available in narrative form. Individual experience can help consultants decide how a framework should be used in a specific project. This means consulting is not just about formal tools or only about tacit judgement. It uses both. More specifically, Werr and Stjernberg (2003, pp. 895–897) argue that codified knowledge has to be translated into a specific situation by someone with practical understanding. Stored knowledge does not work by itself. It has to be adapted by someone who understands the client situation. The authors also suggest that consultants do not experience articulated and tacit knowledge as separate phenomena. Rather, they are interwoven aspects of practice (Werr & Stjernberg, 2003, p. 897). This point is especially important in consulting, where codification includes structured analyses and writing routines. These processes still depend on professional experience and judgement.

Knowledge-intensive professional work also creates value in more than one direction. Løwendahl et al. (2001, p. 911) argue that professional service firms create value both for clients and for themselves through the development of the firm's knowledge base. Each project produces not only deliverables, but may also increase sector understanding, refine methods, and develop consultant capability. Løwendahl et al. (2001, p. 914) also argue that professionals learn from the clients and projects they work with, and that these projects influence both what they learn and how much they learn. Consulting projects are therefore not just repetitions of earlier projects, but they also create opportunities for learning.

This also affects how productivity in consulting should be understood. Løwendahl et al. (2001, pp. 921–923) describe several types of professional service outputs. In consulting, these include information, market analyses, reports, advice, recommendations, innovations, and implementation assistance. This shows that consulting cannot be reduced to one activity. Some parts are close to document production, analysis, and synthesis. Other parts involve problem definition, mediation, stakeholder consultation, and implementation. GenAI may be easier to use in the first group of activities. The second group still depends more on human judgement and understanding of the client context.
Consulting expertise is therefore relevant beyond technical analysis. According to Mosonyi et al. (2020), consulting scholarship is about more than information; it is also about identity and power. This helps avoid a narrow technical view of consulting. Advice has to be credible, relevant, and implementable. Consultants do not only analyse data, but also define the problem, use their expertise, and work within the power dynamics of the client company. Although good analysis plays a part in consulting success, success also depends on how the analysis is understood, accepted, and implemented.

2.3 Human–AI Collaboration in Knowledge Work

Consulting offers a relevant case for examining human–AI collaboration in knowledge work. It involves applying codified methods to the unique circumstances of each client. Methods, frameworks, and cases all matter, yet they do not generate value on their own. They have to be interpreted and adjusted to match a specific client's situation. As noted in Section 2.2, the consulting firm functions as a knowledge system where codified knowledge is combined with individual experience to adapt to a concrete project (Werr & Stjernberg, 2003, p. 881). GenAI can become useful in helping consultants handle tasks that involve applying such codified knowledge.

One way to understand GenAI use in consulting is through the distinction between automation and augmentation. According to Raisch and Krakowski (2021, p. 192), automation refers to a situation where a machine performs a task instead of a human, while augmentation involves human–machine collaboration. They also note that the distinction between automation and augmentation is not always clear, since the two processes can occur at the same time and even create tensions.

Therefore, GenAI is unlikely to replace the consulting process entirely. Instead, it may change how, and by whom certain tasks are undertaken. GenAI does not replace the work of a consultant, but it may be used for selected tasks in the consulting process. It could assist a consultant in summarising interviews, formulating a report structure, or producing presentation material. It does not, however, solve the consulting problem itself, because the consultant still has to organise and interpret the information in light of the client's business needs. According to Raisch and Krakowski, automation in one task can lead to augmentation in another (Raisch & Krakowski, 2021, p. 197). This means that while one task may be automated, the following task may still require a person to set goals and evaluate the outcome.

Jarrahi's concept of human–AI symbiosis adds another view. He argues that organisational decision-making is shaped by uncertainty, complexity, and equivocality (Jarrahi, 2018, p. 577). AI and humans have different strengths in relation to these conditions. AI is strong in large-scale information processing. Humans remain important when the situation is unclear, uncertain, and dependent on judgement or holistic understanding (Jarrahi, 2018, pp. 580–583). In consulting, GenAI may help with information-heavy tasks. Consultants still have to interpret the findings and decide what they mean for the client. Human–AI symbiosis in consulting should not be seen as delegating decision-making to AI. Rather, it means redistributing some tasks between the consultant and the system. According to Jarrahi, human–AI collaboration depends on the individual's understanding of the opportunities and limitations of AI use (Jarrahi, 2018, pp. 583–584). This means that while the system can assist with analytical, hypothesis-related, and language-related tasks, the consultant has to make sure that the advice is timely, properly framed, and acceptable in the client organisation. In such cases, GenAI can help with preparation, idea generation, and text reformulation.

The role of GenAI in consulting can also be examined through recent research on information practices. Jarrahi et al. examine GenAI as part of information practices in knowledge work, rather than only as a tool for isolated task completion (Jarrahi et al., 2025, pp. 1, 28). In other words, the consultant's role does not consist merely of using the system itself. GenAI can be used as a synthesiser, organiser, brainstormer, reviewer, interpreter, and clarifier (Jarrahi et al., 2025, pp. 13–14). These roles are close to tasks that consultants also perform, such as creating summaries, offering alternative explanations, building arguments, reviewing data, and adapting messages for clients.

It is also important not to see GenAI only as a source of ready-made material. According to Jarrahi et al., human–AI collaboration does not simply mean that humans consume computer-generated output. The human collaborator continues to critique, integrate, contextualise, communicate, and individualise the output (Jarrahi et al., 2025, pp. 22–24). Draft generation may speed up the writing process, but human effort is still needed before the document can be used professionally. This includes checking that the output is correct, relevant, and presented in a suitable form.

Iteration is also part of working with GenAI. According to Jarrahi et al., interaction with AI often proceeds through prompts, adjustments, clarifications, and back-and-forth refinement that gradually adapts the output to the user's need (Jarrahi et al., 2025, p. 26). This can be compared to the consulting process, where problem formulation can arise through client discussions, interviews, and analysis.

AI can also create new work around the system. In their study of algorithmic analysis work, Grønsund and Aanestad found that the introduction of AI did not reduce human involvement. Instead, it created new roles and tasks (Grønsund & Aanestad, 2020, pp. 8–12). For example, human effort was needed to audit outputs against a benchmark and to modify the data or the system when mistakes appeared. In consulting, this may mean checking whether the output reflects sources accurately or whether generated advice can be used with a client.

However, as Grønsund and Aanestad (2020, pp. 11–12) show, auditing AI outputs requires a point of comparison. Consultants can validate AI output when there are clear criteria against which the output can be compared. However, when the task involves strategic interpretation or a new client environment, the process becomes more difficult. Human–AI collaboration therefore requires the consultant to analyse and contextualise system output instead of accepting it at face value.

The study by Vaccaro et al. (2024, pp. 2293–2295) provides an essential qualification to this debate. It demonstrates through meta-analysis that human–AI pairings frequently outperform humans acting independently. Still, they do not necessarily outperform the better-performing option, whether human or AI. This trend is more pronounced in content generation than in decision-making.

The core question, then, is not whether AI replaces human decision-making processes; rather, it is how work-related activities are designed. GenAI can help in the generation, synthesis, review, interpretation, and communication of work outputs. The accountability for judging whether the outcome makes sense ultimately remains with the consultant.

2.4 Productivity in AI-assisted knowledge work

As knowledge work evolves through human–AI collaboration, measuring the productivity of such work also becomes more difficult. Outputs in knowledge-intensive work are often intangible and non-standardised, which makes comparisons across individuals, teams, or projects challenging. According to Ramírez and Nembhard (2004), there is no general metric for measuring productivity among knowledge workers because this involves multiple aspects, including quality, output, cost, efficiency, effectiveness, client satisfaction, innovation, and project success. This becomes more visible when AI is introduced, since a tool might reduce the duration of a task without adding value to the output.

To distinguish productivity from pure efficiency, Palvalin (2019, pp. 211–212) emphasises that productivity involves the connection between output and resource consumption, but also takes quality into account. For AI-assisted tasks, this is important since speed does not guarantee quality, relevance, or utility. Therefore, a consultant can complete tasks quicker, but productivity improves only if the output remains accurate and useful. Time saved in preparing a report or building an initial analytical structure is valuable if it gets used for deeper analysis, more client iterations, or better tailoring of the solution. The same saving brings little if it goes into reviewing, fixing mistakes, and correcting unsupported statements. A team might save two extra hours from a faster first draft, but the saved time disappears if a senior consultant spends the same two hours fixing weak arguments. The important question is how the saved time is used. Óskarsdóttir et al. (2022, pp. 1, 26–27) make a related point in their productivity framework. According to them, productivity also depends on what kind of work is being done and what value it ultimately produces for the individual, the organisation, and others in the wider social system (Óskarsdóttir et al., 2022, p. 27).

Professional services also add complexity to the productivity issue. According to Løwendahl et al., many professional service projects are characterised by high customisation. This makes the task difficult to predefine at the beginning of the process, as additional activities and dependencies appear throughout the project (Løwendahl et al., 2001, p. 923). Furthermore, complexity is accompanied by increased coordination costs and the need for mutual adjustment in complex services (Løwendahl et al., 2001, pp. 923-924). A consulting project can start with addressing one question but shift when interviews or data work reveal that the real problem lies elsewhere. Productivity in such projects depends on smooth execution and adaptation.

Therefore, it is helpful to distinguish between task-level productivity, project-level productivity, and organisational productivity. At the task level, productivity concerns quick gains from such activities as theme extraction from documents, preparation of comparison tables, and reorganising analytical data. Next, project-level productivity concerns the overall project quality, timeliness, and relevance. Finally, organisational productivity relates to learning and developing more effective practices. Thus, any productivity gains must ensure that they fit the analysis logic, meet the team requirements, and satisfy the client. Results from human–AI collaboration studies support this, as it tends to bring stronger gains in content creation than in decision-making tasks (Vaccaro et al., 2024, p. 2293). GenAI can also shift work from generation toward review and contextualisation (Jarrahi et al., 2025, p. 25; Grønsund & Aanestad, 2020, pp. 10–11).

Improvements in productivity can also vary considerably across workers. Brynjolfsson et al. found that a generative AI assistant raised productivity by 14% on average, with 34% gains among novice and low-skilled workers and minimal impact among experienced ones (Brynjolfsson et al., 2023, p. 2). The setting was customer support, not consulting, but the pattern is still relevant. Junior consultants may benefit more directly from AI support in structured tasks, while senior consultants doing judgement-heavy work see smaller or no gains.

Knowledge-work research also shows that productivity depends on more than just technology. Palvalin found that well-being at work, individual work practices, and social environment were more strongly related to productivity than physical or virtual environment (Palvalin, 2019, pp. 215, 218–220). Óskarsdóttir et al. similarly note that many factors influencing knowledge-worker productivity are hidden, interacting, and difficult to manage (Óskarsdóttir et al., 2022, p. 1). Even a technically capable model may produce limited benefits if the team lacks a shared review process, clearly defined roles, and proper integration of AI outputs into the workflow.

Reliability is also part of productivity. Feuerriegel et al. explain that current generative AI models can produce incorrect outputs because they generate the most probable response to a prompt rather than necessarily the correct one (Feuerriegel et al., 2024, p. 117). Huang et al. describe LLM hallucinations as outputs that look credible but are factually unsupported and difficult to detect (Huang et al., 2025, p. 42:2). Incorrect synthesis or unsupported claims can damage both decision quality and client trust. A fast workflow becomes unproductive if it produces material that cannot be relied upon.
Productivity should therefore not be equated with improvements in writing or summarisation, even though these improvements are relevant. They are only part of the wider productivity question. A slide deck produced faster has limited value if it fails to address interview findings or match the client's decision situation. By contrast, even small time savings can be highly productive if they allow testing more hypotheses or sharpening the recommendation.

2.5 Risks, governance and expertise

The potential value of generative AI in professional work should be discussed alongside the risks of its use. One of the significant dangers associated with generative AI is automation bias. Lyell and Coiera (2017, p. 423) define automation bias as overreliance on decision support that reduces users' vigilance in information seeking and processing. Automation bias may appear in two ways: omission errors, where users fail to notice missing or incorrect recommendations, and commission errors, where users trust incorrect automated advice too readily (Lyell & Coiera, 2017, pp. 423–424). In a consulting context, commission errors may be especially problematic because, in many situations, the mistake is difficult to identify. The problem stems from the fact that many consulting and similar knowledge-intensive tasks do not have a simple binary correct answer. Instead, correctness is often interpretive.

As demonstrated by multiple empirical studies, users interact with AI algorithms in different ways. According to Dietvorst et al. (2015), algorithm aversion refers to people becoming reluctant to use algorithms after seeing them make mistakes, even when their performance is better than human decision-making in many cases. At the same time, according to Logg et al. (2019), some users increase their reliance on automated advice and see algorithms as more authoritative sources than other people. These phenomena can be regarded as complementary, and each of them can play an important role depending on the context. Thus, the problem of calibrated trust is very challenging when dealing with knowledge-intensive tasks. Automation bias is especially relevant because Lyell and Coiera (2017, p. 426) state that it tends to occur in complex cognitive activities involving hard-to-verify processes.

Lebovitz et al. (2021) deepen this concern. The authors conducted a field study involving the evaluation of multiple AI-driven systems in a hospital environment. Several systems had a high degree of accuracy in a technical sense. However, the researchers observed that these systems were inadequate for solving real-life tasks because the "ground truth" they were based on did not reflect the richer know-how of professionals in situ (Lebovitz et al., 2021, p. 1501). Furthermore, as the authors note, managers experienced significant difficulties when working with labelled data used in these algorithms, including disagreement among experts about which approach was more valid (Lebovitz et al., 2021, p. 1511).

In this context, generative AI faces a specific challenge. As was described in Section 2.4, LLMs may produce what Huang et al. (2025, p. 42:2) term "hallucinations": plausible statements that lack factual support but seem to be correct and coherent nonetheless. Such claims are often human-like and convincing enough to go unnoticed. For consultants, this phenomenon means that they may receive a draft that contains inaccurate, exaggerated, or misleading content without noticing the flaws. Fabricated data, misleading statements, invented references, unjustified generalisations, and similar errors are just as much of a risk as a straightforward mistake in data analysis.

Lee et al. (2025) provide survey evidence on this shift. The survey included 319 knowledge workers. According to the results, respondents who felt more confident about information provided by GenAI were less likely to engage in critical thinking. At the same time, those who had higher confidence in themselves tended to think more critically (Lee et al., 2025, p. 1). The authors conclude that AI-assisted knowledge work reallocates cognitive effort, shifting its focus towards response integration and information verification (Lee et al., 2025, p. 1). This finding supports the idea discussed above that generative AI does not eliminate mental effort but redirects it. Some tasks become easier, especially drafting and gathering information, while others require more effort from professionals, especially reviewing GenAI outputs, selecting useful data, and integrating the information into final outputs. While the exact impact of this shift is difficult to assess with the current knowledge base, it could have significant implications for the consulting industry.

In addition, generative AI is linked to several legal challenges. Specifically, Ruschemeier (2025, p. 1) observes that AI tools raise concerns related to data protection regulation, including GDPR. The main point is that large language models use sophisticated methods for analysing vast datasets. As a result, both data processing and outputs based on inputted information may involve personal data (Ruschemeier, 2025, p. 3). Furthermore, as noted by Ruschemeier (2025, p. 12), some legal provisions related to GenAI use are difficult to implement in practice. Large datasets, unclear data flows, and difficulty removing or modifying certain items mean that not all rights guaranteed to data subjects can be easily exercised. Since consulting activities involve various types of sensitive information, such as internal reports, strategies, workforce-related statistics, and financial data, these circumstances imply that AI use in consulting should be carefully governed, taking into account existing requirements for the processing of different types of information.

At the same time, legal compliance is only one aspect of GenAI governance in professional services. According to Trincado-Munoz et al. (2025, p. 465), the culture, education, and norms of professionals contribute significantly to maintaining trust and preserving the relationship with clients. Specifically, uncoordinated adoption of AI in professional activities may harm this relationship and weaken the concept of professional fiduciary responsibility (Trincado-Munoz et al., 2025, pp. 465–466). The market-for-lemons argument formulated by the authors means that clients may find it difficult to differentiate professional advice from freely accessible information that might look equally valid. Overall, AI governance should not be considered only in terms of legal obligations. It is also important to ensure that AI does not undermine professional practices. Consulting firms should therefore establish policies on AI use, review standards, documentation of GenAI use, escalation procedures, and similar issues.

One important concern in the use of AI in professional services is its effect on competence formation. As noted by Lebovitz et al. (2021, p. 1517), much professional work is based on know-how that is difficult to formalise. In addition, Lee et al. (2025, p. 15) argue that GenAI should be used in a way that supports verification, response integration, task stewardship, and the preservation of problem-solving abilities. Furthermore, Trincado-Munoz et al. (2025, p. 465) suggest that expertise is supported by professional culture, education, and norms.

Taken together, these sources imply that the adoption of AI could affect the professional development of consultants. The problem is that, if professionals do not actively use GenAI outputs for building skills and competencies, the technology is unlikely to support knowledge and expertise development. At the same time, as explained above, AI use increases the workload related to information verification and analysis. This process may weaken competence formation if it is not supported by feedback, supervision, and active review.

2.6 Theoretical synthesis and research gap

Sections 2.1–2.5 frame management consulting as professional service knowledge work, where value is created through analysis, interpretation, and client-specific problem solving. Research on consulting firms as knowledge systems, professional service firms, and knowledge-intensive business services supports this perspective (Werr & Stjernberg, 2003, p. 881; Løwendahl et al., 2001, p. 911; Bettencourt et al., 2002, p. 101). This is crucial, as generative AI in management consulting is not considered a generic productivity tool. Instead, its effects depend on validation and contextualisation required to transform outputs into something of professional value to clients.

Within this context, generative AI is neither seen an autonomous replacement of human labour nor a sure tool in increasing productivity. Literature on collaboration between humans and AI suggests that generative AI might enhance structured information processing while leaving human judgement in place in uncertain and ambiguous situations (Jarrahi, 2018, pp. 577, 580–583; Raisch & Krakowski, 2021, p. 192). The tasks of many advisors involve both of these components. Hence, generative AI may change how work is done, in particular affecting generation, synthesis, and drafting activities. These tasks are valuable because of human work performed afterwards: interpretation, contextualisation, and taking responsibility.

This means that time spent on a given task is not always equivalent to productivity in consulting work. As mentioned earlier in Sections 2.1 and 2.4, consulting productivity involves quality and professionalism beyond mere speed (Palvalin, 2019, pp. 211–212; Óskarsdóttir et al., 2022, pp. 1, 26–27; Løwendahl et al., 2001, p. 911). Saving time at the stage of drafting may be absorbed later by verification and rework. Instead of asking only about time savings, the more relevant question is whether they lead to better decision-making, improved output quality, or lower project risk.

One recurring mechanism across the literature is the shift from generating outputs to validating them. The outputs generated by generative AI could help advisors work faster with idea expansion and drafting. Humans then validate those outputs, integrate them into a coherent account of the problem at hand, and turn them into professional work (Jarrahi et al., 2025, pp. 22–23; Lee et al., 2025, pp. 15–16; Grønsund & Aanestad, 2020, pp. 10–11). This shift comes with real challenges. Validation requires professional judgement based on experience, standards, and tacit skills that formal accuracy measures do not capture well (Lebovitz et al., 2021, p. 1501). Hence, increased workload at the stage of validation may reduce the value of time saved during drafting.

As mentioned earlier, risk and governance are not merely constraints to generative AI but integral parts of its functioning. Overreliance on AI, hallucinations, confidentiality risks, and lack of effective governance hinder proper implementation and use of the outputs generated through AI in professional practice (Lyell & Coiera, 2017, pp. 423–426; Huang et al., 2025, p. 42:2; Ruschemeier, 2025, pp. 1–3, 10–11). For consulting organisations, productivity gains achieved through AI are therefore conditional and depend on their ability to govern the technology properly. Despite this conceptual foundation, important gaps remain. First, the reviewed literature provides limited direct evidence on how generative AI affects entire management consulting workflows rather than bounded tasks. Although some of the reviewed literature touches upon this issue, most studies focus on moments of human–AI collaboration in drafting, prediction, or validation stages. Dell'Acqua et al. (2023) partly clarify the task-level issue by showing that performance may improve in some tasks and worsen in others, depending on where the task falls relative to the model's capability frontier. In other words, strong task performance does not guarantee workflow performance. Second, even though time spent on a task is not necessarily a proxy for productivity in knowledge work, consulting-specific research still provides limited explanation of how time savings become consulting productivity. Third, the influence of generative AI on the development of expertise in the field deserves more attention. As discussed earlier, AI may undermine early-stage reasoning, a vital part of the learning process. At the same time, generative AI paired with feedback, explanations, and reviewing could aid learning (Lee et al., 2025, p. 15; Lebovitz et al., 2021, p. 1517; Trincado-Munoz et al., 2025, p. 465). This is a research gap worth examining specifically in the context of consulting.

Moreover, there is another difference between experimental data and real-world consulting practice, namely a difference in timeframes. In experiments, researchers usually measure the time required to complete bounded tasks. In consulting practice, a project comprises a sequence of interrelated tasks. For these reasons, this thesis addresses a relevant research gap. The existing body of knowledge explains why consulting is an appropriate domain to study generative AI and what kind of productivity consulting involves. What remains underexplained is how generative AI affects management consulting work processes and how task-level gains develop into realised productivity.

3 Methodology

This thesis applies an integrative literature review approach to build an analytical synthesis of secondary empirical evidence. No primary data is collected. The study analyses published empirical studies, adoption and survey evidence, consulting industry reports, and selected conceptual literature on GenAI, management consulting work processes, and productivity.

This literature review approach is suitable due to the fragmented nature of GenAI research to date. The goal here is not a literature review in the sense of summarising all GenAI-related research findings. Rather, the thesis seeks to analyse GenAI's impact on consulting and knowledge-work productivity. As Snyder (2019, pp. 333–334) argues, a literature review can be understood as a way of collecting and synthesising previous research in a field or discipline. Similarly, Whittemore and Knafl (2005, pp. 546–547) state that an integrative literature review can synthesise information not only from experimental and non-experimental studies but also from other sources of evidence such as theoretical literature.

Such a review aims to analyse both the performance outcomes and mechanisms of GenAI integration. In particular, the study considers how GenAI impacts the work processes and why findings vary across tasks and knowledge settings. No single type of evidence is sufficient for these aims. While experimental studies can show performance effects under certain conditions, survey evidence can indicate adoption rates and usage of technologies in working life. Consulting and industry reports can offer explanations for value creation and organisational challenges at the level of business processes. The study should therefore be understood as an analytical and integrative synthesis of existing evidence.

The integrative synthesis is organised through a three-part evidence structure. This is not a separate research method, but a way to compare different evidence types. The first evidence category consists of experimental and quasi-experimental studies, such as Dell'Acqua et al. (2023), Noy and Zhang (2023), Brynjolfsson et al. (2023), and Hao et al. (2024), and provides insight into performance, quality, productivity, or decision support at the task level. The second evidence category consists of adoption and survey studies, with Bick et al. (2025) as the main source for assessing the extent to which GenAI is adopted and what share of work is affected. Finally, consulting and industry reports, such as McKinsey & Company (2023), Accenture (2024), and Deloitte (2024), contribute valuable information about organisational potential, challenges of scale, value creation, and business process considerations.

3.1 Data collection

The material was collected from Google Scholar, selected Tritonia library databases, and the official websites of major consulting firms. Academic material was sourced from business and management databases, multidisciplinary citation databases, publisher platforms, information systems databases, and legal and institutional sources when needed. Key Tritonia databases in this category include Business Source Ultimate, ABI/INFORM Complete, Scopus, Web of Science, ScienceDirect, Springer Nature Link, SAGE Journals, Wiley Online Library, Emerald Journals, ACM Digital Library, AIS eLibrary, PubMed, OECD iLibrary, and EU legal databases such as EU Law and Publications and EUR-Lex. Reports from the consulting industry were gathered from official firm websites. Searches were conducted in February–April 2026 and further updates were made during the writing process for sources appearing to provide directly relevant evidence. Search terms included "generative AI", "consulting productivity", "knowledge work AI", "AI productivity", "management consulting AI", and "generative AI knowledge work". Search terms were applied separately or in combination.

The searching process was based on the logic of an integrative literature review. Snyder (2019, p. 336) notes that literature reviews should clarify their search terms, databases, and inclusion and exclusion criteria, as these factors influence the quality and credibility of the process. Similarly, Whittemore and Knafl (2005, pp. 548–549) note that an integrative literature review needs to document the steps of the literature search, including search terms, databases, additional search strategies, and the criteria for including relevant sources. Following this recommendation, the thesis does not claim to find all potentially relevant sources concerning GenAI and consulting but tries to gather and document a focused and relevant body of material.

The empirical search focused mainly on sources published between 2022 and 2026 due to the technological context of the study. The current wave of GenAI entered everyday workplace use only after large language model tools became widely accessible in late 2022. Literature from before 2022 may provide useful theoretical background, but it does not capture the work practices, technologies, and adoption phases most relevant for this study. The empirical search therefore focused on a period in which GenAI became relevant to knowledge work and professional services.

Sources were initially screened through titles, abstracts, and executive summaries. Those passing initial screening were read in full. Sources were selected using three criteria. First, the source had to be relevant to GenAI and to work processes, productivity, knowledge work, consulting, and professional services more broadly. Second, the source had to represent the current phase of GenAI development unless it was used for theoretical purposes. Third, the source had to reach a certain threshold of quality. This means that at least one of the following factors was sufficiently transparent in the source to allow critical review: methodology, data basis, authorship, publication type, or institutional provenance. Empirical studies, high-quality working papers, survey-based evidence, and sources with clear methodology were preferred for the thesis. Working papers were not ruled out automatically due to the evolving nature of the topic, as some key studies have not yet been published as journal articles.

An additional criterion related to task similarity to consulting work. Research literature about GenAI and management consulting specifically is currently limited. Therefore, this thesis also drew on literature related to analogue work that is close to consulting in its task characteristics, such as information synthesis, professional writing, information evaluation, decision support, checking and validation work, and client-oriented communication. Noy and Zhang's paper was included on this basis because its occupation-specific writing tasks included consultants and involved activities similar to consulting work (Noy & Zhang, 2023, p. 2). Results from these kinds of sources were carefully evaluated and were considered analogue rather than direct evidence.

The material was classified according to its relevance to management consulting. Direct consulting evidence includes studies that directly examine consulting, consulting firms, or tasks performed by consulting professionals. Consulting-relevant analogue evidence covers adjacent knowledge-work settings not specific to consulting but involving similar tasks, such as synthesis, professional writing, information evaluation, decision support, checking, and client communication. This distinction helped avoid overstating findings from non-consulting contexts.

The consulting reports were included for a specific reason rather than as substitutes for academic literature. The McKinsey & Company (2023) report was selected because it models GenAI's productivity potential across different work activities and occupational groups. Accenture (2024) was chosen for its combination of executive survey data, earnings call analysis, company-level financial data, and GenAI productivity modelling. The Nordic report by Deloitte (2024) was included for its survey-based evidence on GenAI adoption, scaling barriers, governance issues, and value realisation in Nordic firms. These reports were treated as indicative sources rather than as causal proof of productivity effects.

Sources were excluded when they discussed AI only in a general strategic context with no connection to work processes or productivity, or when their content could not be verified from an original publication or official report. The final material consisted of experimental and quasi-experimental studies, adoption and survey-based sources, consulting industry reports, governance-oriented sources, and conceptual literature on knowledge work, professional service firms, human–AI collaboration, and productivity.

3.2 Analysis method

The selected material was analysed through thematic analysis within an integrative synthesis. The sources were read several times, recurring findings were identified, and results were grouped into broader themes in response to the research question. Thematic analysis can be defined as a method for identifying, analysing and reporting patterns of meaning within data (Braun & Clarke, 2006, p. 79). In this thesis, thematic analysis was combined with the logic of an integrative review. This meant reducing the material, displaying it in a comparable form, comparing evidence types, and drawing and checking conclusions before synthesising the findings across the included evidence (Whittemore & Knafl, 2005, pp. 550–551).

It was not sufficient to determine whether the influence of generative AI on consulting was positive or negative. There was also a need to identify the task type, the type of evidence provided, and the productivity indicator considered. This was necessary because productivity means different things across sources. Some studies focus on task speed, others on accuracy, and others still on broader business value. Keeping these distinctions clear was necessary, because treating them as the same thing would have distorted the analysis.

Categorisation was derived from the research question as well as from the structure of consulting work, covering analysis and information processing, reporting and communication, and decision-making and problem-solving. There was therefore a deductive basis for the initial coding. Repeated reading of the material then revealed an inductive layer, with categories being refined to better match the emphasis expressed in the sources. As stated by Braun and Clarke (2006, pp. 83–84), two approaches to thematic analysis are possible, and themes can be either generated through the data or developed beforehand through an analytical orientation.

Two levels of coding were applied. The first level focused on organising material based on elements of consulting work that recur across the sources. The second level referred to cross-cutting analytical themes, such as task speed, verification and review activities, automation bias, adoption conditions, usefulness for consulting clients, and the movement from time savings to broader consulting productivity. These second-order themes became the foundation of the empirical chapter. The purpose was not to provide a separate summary of each source. Rather, the aim was to find patterns across different types of evidence.

The three-part evidence structure provided a framework for comparing different types of evidence. Experimental evidence carried the most weight when task-level effects were considered. Survey and adoption evidence helped estimate the actual extent of GenAI use. Evidence from industry reports was used mainly for organisational context and value claims. A more critical perspective was used when the reports discussed potential rather than realised benefits. Some types of evidence were not considered equally strong. Their role depended on the type of claim being made.

Productivity figures were interpreted carefully. The aim was not to average all findings into one overall number. Instead, the focus was on what the findings meant in their own context. Completing tasks faster does not necessarily mean that consulting work becomes more productive. The key question was whether the reported improvement could support better analysis, clearer communication, better decision-making, lower risk, or shorter project lead time.

While using multiple sources of evidence, the analysis remained within the consulting scope. Every pattern found in the literature was placed into the context of a consulting process. This helped relate findings from other knowledge-work contexts to tasks such as market research, information synthesis, slide creation, client communication, hypothesis testing, and decision support. Evidence from other contexts was applied cautiously and treated as consulting-relevant only when the task structure was sufficiently similar.

3.3 Limitations of the study

The main limitation of this work is its reliance on secondary data. No interviews or surveys were conducted with consultants, and no case studies were conducted with consulting companies. The thesis is based on published literature and is unable to assess whether GenAI has been deployed inside a specific company or how its deployment impacts consultants. This is a normal limitation in literature-based research, yet this is especially relevant to the research at hand since consulting work depends heavily on organisational specifics and client relationships, information about which may be difficult to access via published literature alone. This is why the quality of this kind of review depends heavily on how transparently the material is selected, analysed, synthesised, and presented (Snyder, 2019, pp. 336–337).

Second, the quality of the evidence base must be addressed. There are few direct studies of GenAI use in management consulting, which means that part of the analysis depends on broader knowledge-work results. This choice is appropriate given the similarities between tasks, yet this limits generalisability. Not all relevant sources exist as published journal articles, and some of the most relevant ones remain working papers. Despite this, Dell'Acqua et al. (2023) and Noy and Zhang (2023) were included due to their methodological rigour and direct relevance to the topic. There is also a high degree of heterogeneity across studies in terms of tasks, populations, and measures, making quantitative synthesis difficult.

The third limitation pertains to consulting firm reports. While these reports allow for assessing adoption patterns, organisational change, and managerial views on productivity, they do not function as unbiased indicators of realised productivity. Reports by McKinsey & Company (2023) and Accenture (2024) concern modelled or potential effects, while Deloitte's Nordic report (2024) presents managerial views and survey data. Some optimistic framing can therefore be expected. Industry reports have been interpreted through the lens of scaling logic and organisational framing rather than treated as indicators of realised consulting productivity.

Finally, generalisability remains limited by geography and sector. Bick et al. (2025) provide evidence on U.S. GenAI adoption, while Deloitte (2024) focuses on Nordic organisations. Experimental studies also necessarily simplify real consulting projects into bounded tasks. Though these limitations have been pointed out, they do not affect the applicability of this method. The results should be considered situational and analysed with care rather than treated as generally applicable findings.

4 Empirical evidence synthesis: GenAI in management consulting work processes and productivity

This chapter evaluates findings in the existing evidence concerning the application of GenAI in management consulting and related knowledge-intensive work. The aim is not to treat all source types as equally strong, but rather to distinguish the contribution each type of evidence can make to understanding GenAI's impact in consulting.
Experimental studies are used mainly to understand GenAI's effects in bounded tasks. Adoption studies and industry reports are used to examine how widely GenAI is used and what organisational conditions influence its value.

The empirical basis consists of three evidence groups. The first group includes experimental studies using consulting-like tasks, such as Dell'Acqua et al. (2023). The second group includes survey-based evidence on adoption and usage intensity in professional settings, such as Bick et al. (2025). The third group consists of consulting industry reports on productivity and process implications, including McKinsey & Company (2023), Accenture (2024), and Deloitte (2024). The chapter follows the analytical structure explained in the methodology section and uses these sources together without relying too heavily on any single data source.

Experiments capture bounded causal effects, surveys report actual usage levels, and industry reports highlight organisational potential and implementation difficulties. Since Dell'Acqua et al. and Candelon et al. (2023) draw on the same BCG experiment, they are not treated as two completely separate studies. The Dell'Acqua et al. paper is used as the main academic source, while the Candelon et al. article is used to show how the same results can be interpreted from a management consulting perspective. The table below helps to understand how different kinds of sources are used in this thesis.

Table 1 Evidence types used in the thesis

Evidence type	Example sources	Role in the thesis	Main limitation
Direct consulting evidence	Dell'Acqua et al. (2023); Candelon et al. (2023)	Evidence on GenAI performance in consulting-style tasks and the jagged technological frontier.	The evidence is based on bounded experimental tasks, not full consulting projects or long-term client engagements.
Consulting-relevant analogue evidence	Noy and Zhang (2023); Brynjolfsson et al. (2023); Doshi and Hauser (2024); Hao et al. (2024); Lebovitz et al. (2021)	Supports analysis of writing, decision support, validation, and worker differences.	Not directly focused on management consulting
Adoption and survey evidence	Bick et al. (2025); Deloitte (2024; 2025); Lee et al. (2025)	Examines GenAI adoption, usage intensity, and implementation barriers.	Survey evidence is often self-reported
Consulting industry reports	McKinsey & Company (2023); Accenture (2024); Deloitte (2024; 2025)	Discusses productivity potential, workflow redesign, and scaling challenges.	Often based on modelling and executive perceptions.
Conceptual and theoretical literature	von Nordenflycht (2010); Raisch & Krakowski (2021)	Provides the conceptual foundation for consulting, knowledge work, and human-AI collaboration.	These sources explain concepts and mechanisms, but they do not provide direct empirical evidence
Governance, risk, and legal sources	EDPS (2025); Huang et al. (2025); Ruschemeier (2025)	Supports the analysis of automation bias, hallucinations, validation work, data protection, confidentiality, professional trust, and governance requirements.	Mostly conceptual or legal rather than productivity focused.

The reliability of evidence can differ based on the kind of claim it underpins. Experimental evidence is used to interpret task-level effects. Adoption evidence is useful for measuring the intensity of GenAI adoption by professionals. Consulting industry reports serve as evidence for discussing modelled potential and organisational limitations rather than being viewed as direct evidence of realised productivity. However, it needs to be recognised that success achieved in a controlled setting cannot be guaranteed in real consulting projects or at the organisational level. The main issue in this chapter is the gap between controlled experiments and real consulting work.

One of the central observations in this chapter is that the impact of GenAI on performance is highly context-dependent and specific to the task at hand. This means that task-level improvements themselves cannot be treated as evidence of general productivity increases in consulting. According to Dell'Acqua et al. (2023), the speed and quality of BCG consultants' work increased when they used AI tools in some tasks. Nevertheless, adoption data show that these tools continue to affect only a relatively small proportion of total working time (Bick et al., 2025). Even when improvements occur at the level of individual tasks, shorter completion times and better recommendations require wider adoption and changes in how work is evaluated. This means that some tasks are within the current capability of AI, but other relatively similar tasks are outside its capabilities. According to Dell'Acqua et al., in tasks inside this frontier, consultants using AI completed more tasks, worked faster and produced outputs with higher quality. On the other hand, when the task was deliberately set outside the frontier, consultants who relied on AI were less likely to reach the correct conclusion. Candelon et al. (2023) report a parallel result in the BCG article: GPT-4 improved performance in creative product innovation, but reduced performance in business problem-solving. These findings imply that the same tool can be useful in one part of consulting work, yet risky in another. Productivity, as discussed in Section 2.4, is not only about time. What matters in the context of GenAI is whether the technology can help make better decisions, create stronger outputs, reduce risk, and deliver more value to the client.

4.1 AI in analytical work and early-stage structuring

It appears that GenAI is best utilised in cases where the task can be expressed as a natural-language problem, and where the creation of a good early version already has value. These include such actions as arranging documents, forming the first hypothesis, considering alternatives, or creating an initial analytical framework. Such activities are common early steps in consulting, but do not cover all aspects of consulting analysis.

Of the current evidence, the field experiment of Dell'Acqua et al. is the closest direct empirical source, involving 758 consultants working on two types of tasks: a creative product-innovation experiment and a business problem-solving experiment (Dell'Acqua et al., 2023, pp. 5–7). GenAI use contributed to higher quality and speed in tasks inside the model frontier. Quality scores from human judges increased by 38.0% in the GPT-only group and 42.5% in the GPT+Overview group (Dell'Acqua et al., 2023, p. 10). The time to the completion of the final synthesis task decreased by 27.63% and 22.5%, respectively (Dell'Acqua et al., 2023, p. 12). This is a large gain, but in bounded experimental tasks, not entire client projects.

Candelon et al. (2023) report on a similar discrepancy between creative product innovation and business problem solving in practice. The results indicate that GenAI appears more effective when it comes to generating ideas, building an argument, and persuasive writing, but less efficient in business problem solving, where the right conclusion depends on the correct synthesis of facts. This finding implies that AI may have a stronger positive effect during early material preparation than during the decision-making stage.

In terms of workflow, consulting projects typically start with the initial analysis stage, involving the review of client documents, analysis of the interview notes, organising market and competitor information, and developing the first draft of an issue tree. If this process becomes faster, more time can theoretically be spent on assumption testing, risk evaluation, and the establishment of decision criteria before giving the final recommendation. These activities also make up the second step during which the credibility of the recommendation will be established.

As the jagged frontier concept suggests, GenAI provides benefits in some tasks which could be offset by performance declines in others. According to the study, use of GenAI in activities selected to be outside of the frontier led to 19 percentage points lower likelihood of producing correct solutions (Dell’Acqua et al., 2023, p. 14). Likewise, BCG reports significant drops in performance in cases of GenAI users dealing with business problem solving, as the average score in GPT-4 users was 23% lower than for the control group (Candelon et al., 2023, p. 2). This suggests that some tasks carry a higher risk of error when GenAI is used. When there are no quick ways to validate that the solution is correct, or when its firm-specific character is relevant, speed can lead to premature closure of analysis.

This means that GenAI increases analytical throughput, but not necessarily analytical accuracy. For consulting analysis, it implies the following: less effort will be invested into creating an initial version of material, while more effort will be put into checking whether the created version is correct. This corresponds to the observations of knowledge workers that, in AI-assisted tasks, the focus changes from ideation to information verification and response integration (Lee et al., 2025, pp. 1, 8). The EDPS recommendations follow the same logic: to minimise hallucinations and inaccuracies, manual verification steps should be performed (EDPS, 2025, p. 27).

Adoption data indicates that the task-level advantages do not necessarily get incorporated equally. Currently, GenAI assists between 1% and 7% of total work hours, while reported time savings are about 1.4% of all work hours (Bick et al., 2025, p. 2). Deloitte observes that many businesses continue to operate "at the speed of organizational change, not at the speed of technology" (Deloitte, 2025, p. 3). For consulting projects, it means that effects observed in experimental settings are more likely to emerge within selected task pockets, rather than full project workflows. Overall, the evidence suggests that GenAI plays an important role in analytical tasks, contributing to their throughput, but leaving diagnosis, contextualisation, and final judgement largely unchanged. From this perspective, GenAI appears most useful in speeding up early analytical production. It can help teams create first versions faster. The more valuable part of consulting still lies in checking assumptions, interpreting findings, and deciding what those findings mean for the client organisation.

4.2 AI in professional writing and reporting

In professional writing and reporting, the evidence is more consistent than in decision-making. GenAI seems especially useful in cases when the task includes the production of a first draft, improvement of wording or organisation of the material. In a preregistered experiment involving 444 college-educated professionals, Noy and Zhang report that availability of ChatGPT decreased time needed for writing and increased grade scores by −0.83 SD and +0.45 SD respectively (Noy & Zhang, 2023, p. 4). The experiment involved occupation-specific tasks resembling real-world professional activities, and the occupations of the respondents included consultants (Noy & Zhang, 2023, p. 2). Notably, the performance gap between stronger and weaker performers narrowed when ChatGPT was available (Noy & Zhang, 2023, p. 6).

Understanding the mechanisms by which GenAI assists in improving writing helps interpret this finding. Specifically, Noy and Zhang discovered that the use of ChatGPT reduced rough-drafting time by more than half. However, editing time more than doubled (Noy & Zhang, 2023, p. 7). Moreover, Noy and Zhang note that 68% of the participants who received a prompt to use ChatGPT submitted the text produced by the tool without further editing (Noy & Zhang, 2023, p. 5). Thus, the available evidence favours the interpretation of improvement in the first-drafting stage of the writing activity rather than complementarity in analysing the problem. When considering consulting activities, the result implies the need for a more modest proposition about the potential of GenAI to generate a first draft of any work. This activity is still subject to editing, validation, and adaptation to the client's needs.

In consulting, written products play a key role in making analysis accessible to clients. Slide decks, memos, business cases, and executive summaries carry analytical work into decisions. With faster generation of a first draft, freed time may help focus on assessing argument logic, adapting the document to the client's context, and clarifying assumptions. However, the product remains of limited use if it fails to support the client's understanding and decision-making.

This improvement in writing can itself become a problem to solve. The findings reported by Doshi and Hauser demonstrate this idea through the experiments conducted outside the consulting area. Namely, the authors found that outputs created with AI-generated ideas became more original by 5.4% and 8.1%, and more useful by 3.7% and 9.0% respectively (Doshi & Hauser, 2024, p. 3). At the same time, GenAI writing led to the increase in similarity of the documents created and anchoring them to the initial ideas generated by the technology (Doshi & Hauser, 2024, pp. 4–5). Moreover, Doshi and Hauser identified an "ownership penalty" of at least 25% when a person was informed about GenAI ideas (Doshi & Hauser, 2024, p. 6).

Findings related to GenAI adoption in enterprises support the proposed hypothesis that these advantages are unevenly realised. According to Deloitte's research of GenAI adoption within Nordic countries, enterprises face obstacles in scaling usage of GenAI and introducing risk management measures (Deloitte, 2024, pp. 11–12, 15–16). Writing and reporting require not only the model but also approved tools, data policy, and review. Otherwise, GenAI writing would remain a convenient personal solution for the writer rather than a reliable part of client delivery. The main implication is that GenAI can reduce the time needed to prepare a first draft. It does not replace consultant judgement. Reporting becomes more productive only when faster drafting leads to clearer, better checked, and more client-specific communication.

4.3 AI in decision-making and problem-solving

There is strong evidence indicating that the parts of consulting work where value is highest are also where GenAI carries the most risk. Dell'Acqua et al. report that the correctness of responses drops for a task set beyond the frontier. When completing the outside-the-frontier task, the control group answered correctly around 84.5% of the time, whereas the AI groups managed only 60% and 70% respectively, corresponding to an average drop of 19 percentage points (Dell'Acqua et al., 2023, p. 14). According to

Candelon et al., the use of GPT-4 results in a 23% drop in effectiveness when working through business problems (Candelon et al., 2023, p. 2). However, this is not about the usefulness of the tool itself; the main challenge is that the output can seem useful and steer users to the wrong solution.

One important detail is that reduced correctness did not reduce the perceived quality of recommendations produced with the help of the tool. Dell'Acqua et al. report that in their outside-the-frontier experiment, the quality of recommendation increased by 25.1% in the GPT+Overview group and 17.9% in the GPT-only group regardless of whether the participant's underlying answer was correct (Dell'Acqua et al., 2023, p. 15). In other words, GenAI is capable of enhancing the appearance of answers without enhancing their substance. In consulting, a polished but wrong recommendation may pass through internal review if checking focuses more on presentation than on evidence.

The reasons for this problem include automation bias and overreliance. According to Candelon et al., even training interventions did not mitigate the negative consequences of using AI in the BCG problem-solving scenario; those who underwent training fared even worse on average than their peers (Candelon et al., 2023, p. 9). The authors suggest this may reflect overconfidence resulting from training. Therefore, in addition to prompting, consulting teams must ensure that there is a routine for testing, verifying, and challenging the output of AI tools.

Additionally, GenAI does not linearly increase the speed of decision work. Some of the time saved in producing an answer can move into checking the answer, comparing it with evidence, and resolving inconsistencies. Lee et al. show that confidence in GenAI is negatively associated with critical thinking enactment ($\beta = -0.69$, $p < 0.001$) (Lee et al., 2025, pp. 9–10). Lee et al. (2025, pp. 15–16) show that AI-assisted work still leaves people with checking and steering work, including information verification, response integration, and task stewardship. In consulting, the consultant must still evaluate whether the output is reliable, relevant, and defensible under the client’s circumstances.

Finally, Hao et al. (2024) provide consulting-relevant analogue evidence rather than direct evidence on management consulting projects. Specifically, their quasi-experimental research finds that GenAI can reduce cognitive burden and facilitate structured preparation for decision-making in situations involving unfamiliarity and information overload (Hao et al., 2024, p. 1). On the other hand, the risks associated with overreliance, algorithmic bias, and lack of contextual creativity are also identified (Hao et al., 2024, p. 1). It is recommended that data-driven drafting and content generation be left to GenAI while nuanced understanding, decision-making, and interpersonal processes be reserved for humans (Hao et al., 2024, p. 17). In consulting teams where GenAI produces group output, for example during workshops or steering committee meetings, the output generated by AI may seem neutral because it relies on data. However, the quality of output depends on the quality of the underlying data (Hao et al., 2024, pp. 2–3).

For this reason, consulting organisations should define clear decision gates: what must be verified, who verifies it, and how the verification is carried out. This helps keep GenAI in a supporting role, rather than allowing decision responsibility to shift implicitly to the tool. EDPS guidance follows the same logic. It stresses the need for human intervention and safeguards when AI-generated outputs may influence final decisions (EDPS, 2025, p. 29). For decision-making, the evidence points to a careful use of GenAI. It can help with preparation and option generation. Final judgement should remain with consultants who can evaluate the evidence and take responsibility for the recommendation given to the client. The table below summarises, how generative AI can help the consulting process in specific work phases, and what are the risks on it.

Table 2 GenAI across consulting work phases

Consulting work phase	**Examples of tasks**	**Likely GenAI role**	**Main risk**
Early analysis and structuring	Client document review; interview-note synthesis; first issue tree; initial market overview	First-pass structuring, summarisation, and generation of alternative analytical frames	Weak assumptions, premature closure, or overconfidence in a first structure

Consulting work phase	Examples of tasks	Likely GenAI role	Main risk
Professional writing and reporting	Slide drafts; memos; proposal sections; executive summaries; project updates	Drafting, language improvement, organisation of material, and first version production	Polished but generic output, unsupported claims, or weak adaptation to the client situation
Decision support and problem-solving	Option generation; scenario preparation; preliminary recommendation framing	Preparation of alternatives and supporting material for human judgement	Overreliance, incorrect conclusions, or failure to recognise tasks outside the model frontier
Client communication	Meeting preparation; explanation of findings; workshop material; stakeholder-specific messages	Clarification, reformulation, and adaptation of language for different audiences	Loss of nuance, weak fit with client politics, or communication that appears clear but lacks substance
Review and validation	Source checking; logic checking; contextualisation; evidence comparison; final recommendation review	Limited support through checklists or alternative views, but mainly human responsibility	Review burden shifts to senior consultants, or validation is treated as a superficial step

4.4 GenAI and productivity

It is reasonable here to separate productivity into micro-level and organisational-level effects. The micro-level impact is based on research findings indicating that GenAI can create clear productivity improvements in tasks with defined boundaries. The organisational-level impact depends on the frequency of tool use, process redesign, and whether the extra time is used to create added value for clients. Accenture approaches this issue in a similar way by distinguishing between “knowledge productivity” and “process productivity”, where the first relates to output quality and the second to speed (Accenture, 2024, p. 23).

According to Dell'Acqua et al. (2023), consultants who used AI tools completed 12.2% more tasks, worked 25.1% faster, and the quality of their work increased by no less than 40% on tasks within the model frontier (p. 2). In the field of professional writing, time spent was reduced by 0.83 standard deviations, and grades went up by 0.45 standard deviations (Noy & Zhang, 2023, p. 4). The majority of the effect was experienced by the participants with lower grades (Noy & Zhang, 2023, p. 6). GenAI can serve as a productivity lever in writing, consulting, and support activities. However, performance can decline outside the model frontier, because a fast and incorrect response is still not productive, as it requires corrections, additional time, and creates client risks (Dell'Acqua et al., 2023, p. 14; Candelon et al., 2023, p. 2).

Bick et al. provide an important calibration. GenAI users reported time savings of 5.2%, whereas the average across all employees was only 1.4% (Bick et al., 2025, pp. 19-21). This difference is worth noting, as individual users can see clear time savings, but the overall effect stays modest when usage across the company is uneven. For consulting firms, this means that the tool alone is not enough. The question about how it is adopted and integrated into daily work matters just as much.

In a study on workplace deployment, Brynjolfsson et al. highlight an important nuance about who benefits most from GenAI. Across 5,179 customer support agents, the technology boosted productivity by 14% on average, by 34% for novices and low-skill employees, and negligibly so for high-skill employees (Brynjolfsson et al., 2023, p. 2). Novices and low-skill workers with less than one month of experience gained by 46%, while skilled workers with more than one year of experience did not see an increase in productivity (Brynjolfsson et al., 2023, p. 16). Even though this experiment does not take place in a consulting setting, it still provides relevant analogue evidence: GenAI may benefit less experienced consultants who deal with routine tasks, while more experienced consultants making judgement-based decisions see no or marginal gains.

However, this raises a question of what happens to the saved time. In consulting, freed time may produce client value in at least two ways. First as a quality effect, where the same effort produces a better decision, and second, as a lead time effect, where decisions are made faster while remaining reliable. McKinsey specifically highlights this possibility, stating that efficiency improvements from automation and GenAI depend on the fact that workers spend freed time on alternative tasks whose productivity is at least comparable to existing ones (McKinsey & Company, 2023, p. 45). Additional internal refining will not change the client-side output. Checking sources, producing sharper recommendations, or implementing plans might.

The productivity numbers mentioned in McKinsey and Accenture studies relate to potential estimates and not actual increases. As Accenture states, GenAI may save 12% of work hours and increase output quality by an average of 8.5%, though these are estimates rather than observed firm-level outcomes (Accenture, 2024, pp. 22–23); estimates vary between 10% and 43%, depending on the task. These reports are best read as structured hypotheses about possible gains from implementing GenAI, which may be further evaluated through practical experience.

Deloitte's Nordic survey reveals one of the potential reasons why theoretical estimates of GenAI impact differ from practical results in consulting firms. According to the survey, 35% of Nordic companies had moved more than 30% of their GenAI experiments into full-scale production, whereas 53% of all companies worldwide had done so (Deloitte, 2024, p. 11). Moreover, only 25% of Nordic companies offered GenAI risk training to practitioners, compared to 37% globally (Deloitte, 2024, p. 15). Yet according to Deloitte's later global survey, one-fifth of respondents reported that their organisation’s most advanced GenAI initiative had ROI above 30% (Deloitte, 2025, p. 22). While these numbers are self-reported and refer to best results only, they suggest that significant gains can indeed be achieved once projects go past the experiment stage.

Productive AI usage in consulting appears to depend on two conditions. First, task-level benefits are clearly evidenced in bounded experimental settings, but require broad adoption, process redesign, and strategic choices about how saved time is used. Second, GenAI appears to raise output speed and volume more reliably than it improves correctness. This distinction is what makes the organisational challenge harder than the tool challenge.

4.5 Limitations and conditions of the empirical evidence

Section 2.5 discussed risks related to GenAI at a conceptual level. This section focuses on what the empirical evidence can and cannot show about those risks and productivity conditions. In terms of empirical evidence, the impact of GenAI in consulting depends on certain conditions. First, in addition to lacking comprehensive studies, the findings remain fragmented and capture only specific parts of the broader picture. While many studies provide an effective measure of bounded tasks, other elements, such as entire workflows, interactions with clients, implementation, and realised results, have received much less attention (OECD, 2025, p. 31). Similarly, the adoption of GenAI in consulting is not a straightforward process. As indicated by the later survey by Deloitte, GenAI was accessible to less than 40% of the workforce, and fewer than 60% of those with access used it on a daily basis (Deloitte, 2025, p. 17). Hence, despite any GenAI initiative taken within a particular company, limited access and irregular use may lead to only partial improvements in work processes.

At the task level, the frontier represents a major limitation. As noted before, a consulting project may involve various kinds of activities, some of which would benefit from using GenAI and others which would be hindered. According to experimental evidence, a positive result will be observed within the frontier, whereas outside of it there is reduced performance and correctness (Dell'Acqua et al., 2023, p. 14). As stated by Candelon et al. (2023, p. 2) in their analysis of the results, this effect demonstrates the dichotomy between value creation and value destruction which is characteristic of the consulting profession. Accordingly, risk does not spread evenly in consulting projects; it accumulates at stages requiring intensive judgement under difficult evidentiary conditions. Outside the frontier, human behaviour becomes a separate area of risks. As reported by BCG, despite being explicitly warned about possible errors, participants continued to over-rely on AI-generated output in tasks that went beyond the frontier (Candelon et al., 2023, p. 2). This risk may grow if training leads to an increase in confidence without any improvement in judgement skills (Candelon et al., 2023, p. 9). Lee et al. identified a negative correlation between confidence in GenAI and engagement in critical thinking (Lee et al., 2025, pp. 9–10). In consulting, time pressures and client expectations for quick responses may reinforce this risk of overreliance.

The risk of overreliance in consulting work is tied to the way review activity is structured. GenAI may help save time in generating a first draft but shift it towards correction, verification, and reconciling errors. According to Lee et al. (2025, pp. 1, 8), knowledge workers report verification and response integration as two key forms of critical thinking in GenAI-assisted tasks. Moreover, according to EDPS guidelines (2025, p. 27), manual verification would help avoid hallucinations and inaccuracies in the outputs. Time savings achieved through GenAI are therefore not guaranteed, as they may reappear as increased review burden for senior consultants in consulting projects.

As analogue evidence illustrates, this issue is associated with validation and cannot be solved via formal criteria and surface-level output assessment. According to Lebovitz et al. (2021, p. 1501), systems with high formal accuracy ratings could still face problems in practical implementation if the benchmarks for measuring performance did not consider the context and know-how involved in solving the particular case. This problem is especially relevant for consulting work because its recommendations are often evaluated in terms of framing, client situation, and implementation, among other dimensions which do not appear in benchmark criteria.

Hallucination is an additional complication. Following the terminology of Huang et al. (2025, p. 42:2), one could distinguish factuality hallucinations, when GenAI generates made-up facts, and faithfulness hallucinations, when it fails to follow instructions or adhere to the task context or logic. In other words, the output generated by GenAI could look fine in structure and style yet be insufficiently supported by facts or fail to match the task requirements.

Validation is further complicated by constraints related to data protection. According to EDPS (2025, pp. 20, 36), GenAI could operate with personal data across its entire life cycle, namely in training, testing, validation, input, and output datasets. It is associated with risks like model memorisation, model inversion attacks, prompt injections, and jail-breaking (EDPS, 2025, p. 36). Moreover, Ruschemeier (2025, pp. 10–11) mentions risks of inaccurate GenAI output and difficulty in fixing or deleting such data. Since consulting work involves analysing client projects that include sensitive organisational and personal data even if it seems like a drafting task, the benefit of increased productivity may become difficult to realise.

Data protection risks relate to tighter regulation of GenAI use. According to EDPS, GenAI operation should involve accountability, role definition, documentation, risk assessment, and due diligence of vendors (EDPS, 2025, pp. 7–9). For consulting teams, it translates into stronger requirements for transparency and auditability of client processes and the introduction of policies on tool usage, applicability, and data protection. The method of GenAI application therefore becomes an important compliance question for consulting firms.

These limitations connect back to productivity-related questions. According to the OECD, the major research gaps include understanding how workers handle mistakes made by AI, as well as long-term impacts on business (OECD, 2025, p. 6). If consulting organisations do not recognise which tasks fall outside GenAI capability, time savings may become a source of errors — a risk that is easy to underestimate because GenAI works effectively in drafting and generating outputs.

Across the evidence, GenAI is most beneficial in analysis and reporting tasks but poses a distinct risk profile: automation bias, overreliance, hallucinations, and extra verification effort. Value is realised only when time savings support better decisions and implementable change. In practice, a U-shaped pattern may emerge, as generation becomes faster at first, but verification effort increases until review gates and responsibilities are established. Productivity may become more visible again only after review routines, responsibilities, and tool policies become more stable. This raises a concern for junior consultants. If GenAI replaces routine drafting and basic analysis without proper feedback loops, there is a risk that consultants' critical thinking skills may weaken over time. Lee et al. recommend training that includes output integration, information evaluation, and independent problem-solving (Lee et al., 2025, p. 16), while OECD notes that over-reliance on GenAI may weaken critical thinking and skill development over time (OECD, 2025, pp. 16–17).

5 Discussion and conclusions

Generative AI affects management consulting work processes mainly by making some knowledge-intensive activities faster and easier to start. The clearest impacts of the technology can be observed in situations when consultants prepare an initial version of certain tasks – a summary, a draft slide, an initial issue tree, a market analysis, an early version of a structure, etc. These tasks tend to be common in consulting projects and often precede the formulation of a recommendation to a client. According to the existing empirical evidence, GenAI could potentially facilitate the completion of such preparatory activities if the task is language-related and has enough structure for the system to be able to generate at least a rudimentary draft (Dell'Acqua et al., 2023; Noy & Zhang, 2023).

The research question asks how generative AI affects consulting work processes and under what conditions efficiency at the task level turns into realised consulting productivity. This question has only partial answers. There is some positive evidence that generative AI can help consultants in completing such tasks as analysis, writing, reporting, and certain aspects of preparing for decision-making. Whether it leads to realised productivity depends entirely on subsequent actions. A fast draft can benefit the project if it allows a consulting team to conduct better analysis, communicate more effectively with the client, avoid unnecessary re-work, and save lead times of the process while maintaining quality. On the contrary, a fast but insufficiently sourced and structured draft would lose its advantages in the future.

Analytical work shows good results for generative AI. GenAI can assist in tasks associated with handling information and organising knowledge – summarising documents, generating ideas, finding an optimal structure, and other similar operations. Such conclusions have been made experimentally by Dell'Acqua et al. (2023). At the managerial level, it implies the ability of a team to get an initial draft of an analysis or hypothesis quicker compared to previous practices. It is important to note that at this point of the process, the work would not be done. An analyst should verify the quality of the structure proposed by AI and make sure that it reflects the actual needs of the client.

Writing and reporting also demonstrate considerable evidence. According to Noy and Zhang (2023), using generative AI allowed professionals to complete writing tasks more quickly without harming quality. Given the nature of consulting, these tasks are highly important because they serve as channels of communication with clients. Slideshows, summaries, proposals, and other written materials are used to present conclusions, convince a client of the validity of an idea, and create a basis for further collaboration. As it was previously mentioned, the efficiency of the process does not depend only on speed. A consulting team can write an initial draft faster, but it will not become productive if there are mistakes in it.

Tasks related to decision-making and solving specific problems are not clear enough. Even at the technological frontier, GenAI demonstrates high efficiency in accomplishing some tasks and low efficiency in others (Dell'Acqua et al., 2023). Candelon et al. (2023) have discovered similar tendencies in management consulting, where the AI performed well in tasks requiring creativity and innovation, but poorly when it came to business problem-solving. In management consulting, dealing with problems is especially relevant because of the complexity of client problems and the lack of necessary data for the decision-making process.

Generative AI changes consulting work in part because of the redistribution of effort required to accomplish it. Certain efforts move from first-pass production to the verification of sources, generation of alternatives, elimination of weak logical connections, and deciding how to use the created material. This fact helps to explain the differences between productivity achieved at the task level and the overall project level. While a particular task may become more efficient due to the usage of GenAI, the entire project stays the same if the saved time is not used efficiently.

Productivity at the firm level seems even more conditional. Bick et al. (2025) claim that the adoption of GenAI is already taking place, but its effects on the total number of hours worked by employees remain limited. Reports of such major companies as McKinsey & Company (2023), Accenture (2024), and Deloitte (2024, 2025) demonstrate various possibilities and adoption barriers of GenAI but fail to give any causal evidence on realised productivity changes. Instead, they show that productivity depends greatly on how the tool was integrated, trained, reviewed, and managed in consulting companies.

5.1 Theoretical implications

Management consulting work can therefore be considered an example of knowledge-intensive professional service work, and the findings of this thesis support this view. Management consulting involves not only information processing but also the interpretation of the client's problem, the application of general knowledge to specific cases, and the generation of recommendations credible enough for the client to implement. This explains the uneven impacts of GenAI on various kinds of consulting work: an effective generator of drafts does not necessarily provide similarly valuable results in terms of problem interpretation, priority setting, and the evaluation of what is feasible enough to recommend.

Moreover, the study connects to research on professional service firms and knowledge-intensive business services. According to Løwendahl et al. (2001), such firms create both direct value for clients and indirect value through the development of the firm's own knowledge base. First, the consulting firm provides direct value for its clients by producing valuable outputs. Second, indirect value arises because the consulting firm itself creates and develops knowledge. Both kinds of value might be supported with the help of GenAI. For instance, it might help the team develop clearer analyses or communicate with the client faster. Additionally, the consulting firm might learn to develop better prompts and templates and use them in other cases. However, the translation from using the tool to creating value is not automatic. The same draft can either become project material or cause additional review and evaluation workload.

In this case, the relevant theoretical framework seems to be human–AI collaboration. Jarrahi (2018) suggests that AI and humans have different comparative advantages in knowledge-intensive work processes: AI processes information while humans cope with uncertainty and ambiguity. Therefore, the division of labour in consulting could be pragmatic. The GenAI system could provide the initial version of the analysis that the consultant needs to interpret, understand, evaluate, and transform according to the client's needs. The value of collaboration depends on the consultant's ability to guide the GenAI system and evaluate what it produces.

Finally, the line between automation and augmentation in GenAI use becomes less clear. Raisch and Krakowski (2021) draw a clear distinction between the two. However, this distinction seems problematic with respect to GenAI, since the use of the tool might shift the balance between the amount of labour required in different phases of a task. For example, a simple draft may be created almost automatically, but this draft is followed by a phase that is more demanding in terms of human labour because it requires correction and interpretation.

Regarding productivity, the results support existing viewpoints concerning knowledge-intensive tasks and services. According to Palvalin (2019) and Óskarsdóttir et al. (2022), productivity should take into account several factors. The quality and usefulness of outputs should not be omitted, because increased efficiency could simply mean the faster production of something that was not truly necessary to generate quickly. Therefore, in consulting, the relevant question is whether more rapid production contributes to improved recommendations and decision-making.

5.2 Validation work and the conversion of productivity

The core argument of this thesis comes down to the 'conversion problem.' GenAI can help accelerate the processes of writing, summarising and preliminary analysis, but this does not necessarily mean productivity. The key point here is that AI-generated output has to be assessed, checked against the evidence gathered in the project, adjusted according to the situation in the client organisation, and developed enough for the consultants to back their conclusions. The figure below illustrates how a conversion problem can derail the consulting process if the problem has not been considered.

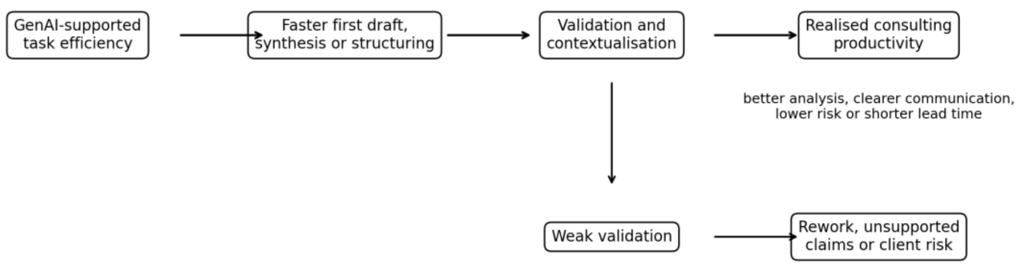

Figure 1 Conversion of task-level efficiency into consulting productivity

Empirically, the evidence points more to a shift from generation to validation than to the elimination of work. While drafting might become faster, the need for editing and reviewing remains. As Noy and Zhang (2023) note, there might be a reduction in drafting time but an increase in editing time. Similarly, Lee et al. (2025) identify that information verification, response integration, and task stewardship may require increased effort after GenAI introduction. From a consulting perspective, a consultant would spend less time crafting the first version of a slide but devote more time to evaluating whether the slide is valid, reasonable, and suitable for presentation to a client.

This validation work should not be overlooked or underestimated. Consulting outputs affect client decisions. A generated market summary, benchmark, or recommendation might sound quite plausible despite relying on poor-quality evidence or reasoning. According to Dell'Acqua et al. (2023, p.14), AI users performed worse in tasks that went beyond the model's capability. In consulting work, it may be more problematic if a recommendation seems convincing but is wrong.

An important aspect here is the presence of a technological frontier, which makes validation work necessary. It may not be safe to assume that GenAI is capable enough to deliver a useful output for every task. Some tasks are easier to iterate because the outputs can be checked quickly or the cost of mistakes is relatively low. Others have higher barriers because checking them requires stronger evidence, or because a mistake may lead to serious consequences.

It is therefore possible that a U-shaped pattern may emerge between GenAI use and productivity. In the early stages, GenAI may help boost productivity because production becomes faster. There may be more options available for consideration without too much additional work. Later on, however, the gains may diminish because more information has to be verified. Consultants may need to spend more time checking information sources, assumptions, and recommendations before using GenAI-produced material.

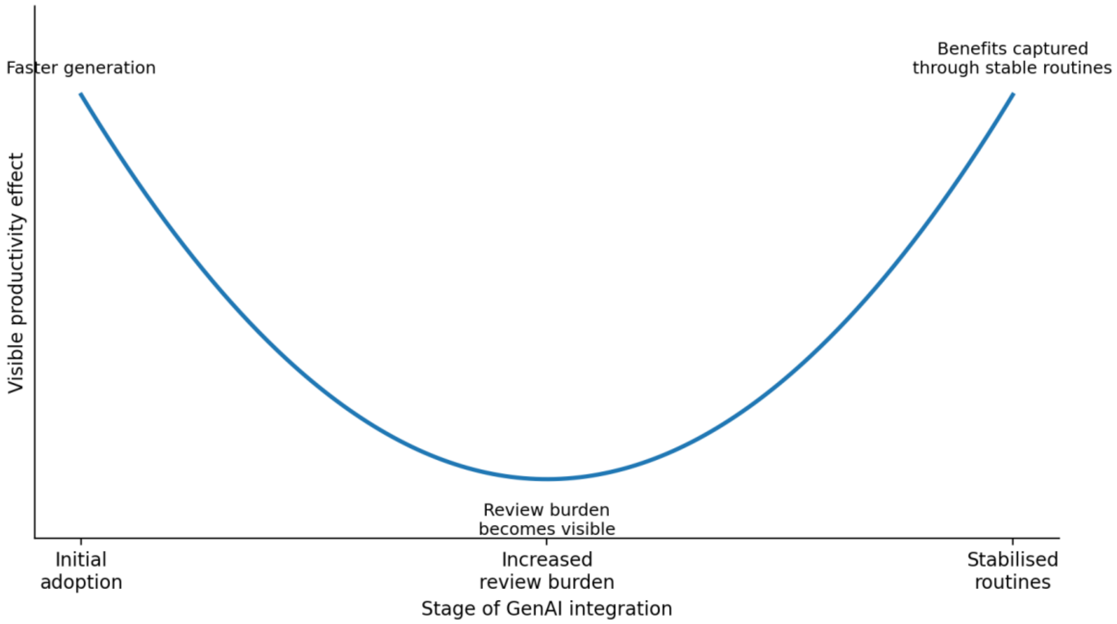

Figure 2 Possible U-shaped pattern of GenAI productivity realisation

Explanation. This diagram presents an interpretation developed in this thesis, not an empirically observed result. It shows how initial efficiency gains from faster production may be offset by increased review work, before productivity gains become more visible again once review routines and responsibilities are established.

At the project level, faster production does not necessarily mean that the whole project takes less time. Productivity improves only if faster production leads to shorter lead time, less rework, stronger recommendations, or better client decision-making. If GenAI merely increases the amount of material the team reviews, there may not be much improvement in productivity.

The need for review connects productivity with risk. Hallucinations are mentioned by Huang et al. (2025, p. 42:2) as one of the critical risks that arise in the case of large language models. Human verification, safeguarding measures, and transparency in GenAI practices are among the key factors emphasised in EDPS guidance on this technology (EDPS, 2025, pp. 7–9, 29). All these aspects cannot be overlooked as separate from productivity. In consulting, weak evidence processing and unsupported claims can damage client trust and increase project risk. A rapid output should not be treated as productive if it creates problems later.

Expert knowledge is another part of the conversion. Lebovitz et al. (2021, p. 1501) show that expert work cannot always be judged through formal accuracy alone. Practical judgement requires knowledge of the context. The recommendation can be technically accurate, yet inappropriate for the client's circumstances. Implementation issues, organisational politics, and stakeholder considerations should also be taken into account. In this regard, GenAI cannot independently determine whether a generated recommendation is applicable in the context of an actual consulting engagement.

5.3 Managerial implications

Companies may need to embed GenAI into the consulting process rather than viewing the technology simply as a way to increase the efficiency of individual consultants. Fragmented use of the technology, such as emailing, summarising, slide-making, and drafting, may benefit some individuals but still lack direction in terms of when and how the technology should be used. It also requires confirmation of the results generated through the technology. From a practical perspective, the key question is to identify which activities can be supported by GenAI.

It would make sense to classify tasks where GenAI is suitable for first-pass generation. These could include document summaries, interview syntheses, slideshows, memos and other internal communication documents, proposal texts, and the generation of options. A higher level of scrutiny should be applied to activities such as final client recommendations, client problem diagnosis, financial or strategic conclusions, and documents used in important decision-making processes. This relates to the uneven nature of the frontier.

Applying the same tool to all activities in the same way is likely to produce inconsistent results.

For validation to succeed, formal checkpoints are needed. Companies need to be clear on which pieces of AI-generated content require validation, by whom, and to what extent. Checking the source documents used in an AI-generated market analysis is one thing; assessing the validity of draft advice is another. This usually requires an expert with a good understanding of the client's operations. Even something as simple as a presentation slide should be analysed in terms of logic, not just proofread. Although creating checkpoints takes time at the beginning, it may prevent a larger amount of rewriting later.

Guidelines on data and tool use should also be defined. Consulting often involves working with highly confidential data, including personal data, methodological knowledge, and business-sensitive information. Sources such as EDPS (2025), OECD (2025), and Ruschemeier (2025) stress privacy, ownership, and accountability issues in AI use. This is especially relevant in consulting, where misuse of client information may seriously affect client trust. This is directly connected to role clarity. The system cannot accept responsibility for its work, so people need to know what roles they are expected to play. Consulting firms must identify who is responsible for checking a particular AI output, whether that person is a junior consultant, project manager, or partner. In addition, the firm has to determine at which point the output should be checked.

Training should focus on verification at least as much as on prompting. Prompting techniques are important because they can help consultants work more efficiently. However, verification is equally critical. Lee et al. (2025) raise concerns about overreliance on the tool and weakened critical thinking. In consulting, where clients rely on the outputs, consultants need to be able to verify facts, recognise weak arguments, and compare the tool's outputs against project evidence.

The training of junior consultants deserves special attention. GenAI can help less experienced consultants structure their work and generate drafts more quickly. Brynjolfsson et al. (2023) suggest that less experienced customer support workers gained more from GenAI than more experienced workers. On the other hand, if junior consultants spend most of their time editing AI-generated output, they may not get enough practice in critical thinking.

When assessing success, reducing time spent is not enough. It is necessary to ask whether the technology helps avoid extra work later, improves the quality of the final output, increases overall project efficiency, or builds reusable know-how. McKinsey & Company (2023) and Accenture (2024) provide useful estimates of the potential benefits that may follow from implementing generative AI. However, despite the useful ideas presented in these and other sources, they do not prove that realised consulting productivity has already improved.

5.4 Limitations and future research

First, the use of secondary sources is the main limitation of this thesis. As mentioned earlier, no surveys, interviews, or case studies were conducted in consulting organisations. Thus, the research primarily analyses consulting-related literature. In addition, it cannot examine the actual implementation of GenAI technology or its impact on consulting practices inside a specific organisation. However, there are also problems related to consulting-specific data. Dell'Acqua et al. (2023) is highly relevant to the topic because it provides evidence on consultants and consulting-like tasks. By contrast, other secondary evidence comes from areas such as professional writing, customer support, medicine, decision support, and related fields. These sources can be used in the analysis only when the task structure is sufficiently similar to consulting.

Industry reports might also be seen as an additional limitation. Reports by McKinsey & Company (2023), Accenture (2024), and Deloitte (2024, 2025) can be useful for identifying potential areas where efficiency might be increased, understanding why organisations struggle to adopt or scale such technologies, and identifying issues related to large-scale implementation. However, many consulting reports are based on modelling and survey responses gathered from organisations.

Due to these limitations, an accurate examination of productivity improvement at the organisational level is difficult to be conducted. Experiments focus on narrowly defined tasks, so the productivity of the whole consulting process cannot be evaluated through them. Surveys can indicate usage figures, but it does not conclusively prove a relationship with increased productivity.

Further studies should explore the application of GenAI in complete consulting processes instead of bounded tasks only. Case studies of individual firms could provide evidence on the use of GenAI in analysis, reporting, client communication, and implementation assistance. Interviews with consultants, project managers, partners, and clients might reveal details about how AI outputs are verified and who takes responsibility for their accuracy. At the project level, it would be useful to explore whether GenAI helps complete tasks more quickly, minimise rework, or increase client satisfaction. The crucial issue remains whether GenAI supports knowledge acquisition or obstructs the experiential learning process that is fundamental to consulting.

5.5 Concluding remarks

Specific parts of management consulting work can be accelerated by generative AI. These include the preliminary creation of summaries, drafts, structures, and material for further analysis. Although these outputs are useful, they do not result in enhanced consulting productivity on their own.

Consulting productivity depends on what happens after output creation. The output needs subsequent analysis and adaptation, and it has to be supported by evidence. Generated material should serve as a basis for decision-making, while consultant judgement still plays a critical role. This also applies to reviewing outputs, managing data responsibility, and developing the workflow. Validation should not be perceived as one more element of consulting; it should be regarded as a necessary part of the process. Therefore, generative AI cannot be considered an autonomous productivity solution for management consulting. Productivity becomes possible through integrating GenAI into consulting processes, validating its outputs, and using it responsibly.

References

Accenture. (2024). *The productivity payoff: Unlocking competitiveness with generative AI*. Accenture. Retrieved 21.04.2026 from https://www.accenture.com/us-en/insights/strategy/productivity-payoff

Bettencourt, L. A., Ostrom, A. L., Brown, S. W., & Roundtree, R. I. (2002). Client co-production in knowledge-intensive business services. *California Management Review, 44*(4), 100–128. https://doi.org/10.2307/41166145

Bick, A., Blandin, A., & Deming, D. J. (2025). *The rapid adoption of generative AI* (Federal Reserve Bank of St. Louis Working Paper No. 2024-027). Federal Reserve Bank of St. Louis. https://doi.org/10.20955/wp.2024.027

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Brynjolfsson, E., Li, D., & Raymond, L. R. (2023). *Generative AI at work* (NBER Working Paper No. 31161). National Bureau of Economic Research. https://www.nber.org/papers/w31161

Candelon, F., Krayer, L., Rajendran, S., & Zuluaga Martínez, D. (2023, September 21). *How people can create—and destroy—value with generative AI*. Boston Consulting Group. Retrieved 19.03.2026 from https://www.bcg.com/publications/2023/how-people-create-and-destroy-value-with-gen-ai

Dell'Acqua, F., McFowland, E., III, Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Krayer, L., Candelon, F., & Lakhani, K. R. (2023). *Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality* (Harvard Business School Technology & Operations Management Unit Working Paper No. 24-013). Social Science Research Network. https://doi.org/10.2139/ssrn.4573321

Deloitte. (2024, November). *Now decides next: Moving from potential to performance in the Nordics. Deloitte's State of Generative AI in the Enterprise,* *quarter three report: Nordic cut*. Deloitte. Retrieved 05.04.2026 from https://nor.deloitte.com/rs/712-CNF-326/images/State%20of%20GenAI%20Nordic%20cut%20Q3%20report.pdf

Deloitte. (2025, January). *Now decides next: Generating a new future. Deloitte's State of Generative AI in the Enterprise, quarter four report*. Deloitte. Retrieved 05.04.2026 from https://www.deloitte.com/nz/en/services/consulting/perspectives/generative-ai-in-enterprise-generating-a-new-future.html

den Hertog, P. (2010). *Managing service innovation: Firm-level dynamic capabilities and policy options* [Doctoral dissertation, University of Amsterdam]. Retrieved 18.03.2026 from https://dare.uva.nl/id/51e96fec-bc8b-4550-96ab-a90a24ee0e77

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General, 144*(1), 114–126. https://doi.org/10.1037/xge0000033

Doshi, A. R., & Hauser, O. P. (2024). Generative AI enhances individual creativity but reduces the collective diversity of novel content. *Science Advances, 10*(28), Article eadn5290. https://doi.org/10.1126/sciadv.adn5290

European Data Protection Supervisor. (2025, October 28). *Generative AI and the EUDPR: Orientations for ensuring data protection compliance when using generative AI systems* (Version 2). European Data Protection Supervisor. Retrieved 04.04.2026 from https://www.edps.europa.eu/system/files/2025-10/25-10_28_revised_genai_orientations_en.pdf

Feuerriegel, S., Hartmann, J., Janiesch, C., & Zschech, P. (2024). Generative AI. *Business & Information Systems Engineering, 66*(1), 111–126. https://doi.org/10.1007/s12599-023-00834-7

Grønsund, T., & Aanestad, M. (2020). Augmenting the algorithm: Emerging human-in-the-loop work configurations. *The Journal of Strategic*

Information Systems, 29(2), Article 101614. https://doi.org/10.1016/j.jsis.2020.101614

Hao, X., Demir, E., & Eyers, D. (2024). Exploring collaborative decision-making: A quasi-experimental study of human and generative AI interaction. *Technology in Society, 78*, Article 102662. https://doi.org/10.1016/j.techsoc.2024.102662

Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., Chen, Q., Peng, W., Feng, X., Qin, B., & Liu, T. (2025). A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. *ACM Transactions on Information Systems, 43(2), Article 42, 1–55*. https://doi.org/10.1145/3703155

Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human–AI symbiosis in organizational decision making. *Business Horizons, 61*(4), 577–586. https://doi.org/10.1016/j.bushor.2018.03.007

Jarrahi, M. H., Li, L., Robinson, A. P., & Meng, S. (2025). Generative AI and the augmentation of information practices in knowledge work. *Behaviour & Information Technology*. https://doi.org/10.1080/0144929X.2025.2551570

Lebovitz, S., Levina, N., & Lifshitz-Assaf, H. (2021). Is AI ground truth really "true"? The dangers of training and evaluating AI tools based on experts' know-what. *MIS Quarterly, 45*(3), 1501–1526. https://doi.org/10.25300/MISQ/2021/16564

Lee, H.-P. H., Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R., & Wilson, N. (2025). The impact of generative AI on critical thinking: Self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery. https://doi.org/10.1145/3706598.3713778

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and*

Human Decision Processes, 151, 90–103. https://doi.org/10.1016/j.obhdp.2018.12.005

Løwendahl, B. R., Revang, Ø., & Fosstenløkken, S. M. (2001). Knowledge and value creation in professional service firms: A framework for analysis. *Human Relations, 54*(7), 911–931. https://doi.org/10.1177/0018726701547006

Lyell, D., & Coiera, E. (2017). Automation bias and verification complexity: A systematic review. *Journal of the American Medical Informatics Association, 24*(2), 423–431. https://doi.org/10.1093/jamia/ocw105

McKinsey & Company. (2023, June). *The economic potential of generative AI: The next productivity frontier*. McKinsey & Company. Retrieved 26.03.2026 from https://www.mckinsey.com/capabilities/tech-and-ai/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier

Mosonyi, S., Empson, L., & Gond, J.-P. (2020). Management consulting: Towards an integrative framework of knowledge, identity, and power. *International Journal of Management Reviews, 22*(2), 120–149. https://doi.org/10.1111/ijmr.12218

Noy, S., & Zhang, W. (2023). *Experimental evidence on the productivity effects of generative artificial intelligence*. Working paper, MIT. https://doi.org/10.2139/ssrn.4375283

OECD. (2025). *The effects of generative AI on productivity, innovation and entrepreneurship* (OECD Artificial Intelligence Papers No. 39). OECD Publishing. Retrieved 17.04.2026 from https://www.oecd.org/en/publications/the-effects-of-generative-ai-on-productivity-innovation-and-entrepreneurship_b21df222-en.html

Óskarsdóttir, H. G., Oddsson, G. V., & Unnthorsson, R. (2022). Towards a holistic framework of knowledge worker productivity. *Administrative Sciences, 12*(2), Article 50. https://doi.org/10.3390/admsci12020050

Palvalin, M. (2019). What matters for knowledge work productivity? *Employee Relations, 41*(1), 209–227. https://doi.org/10.1108/ER-04-2017-0091

Pyöriä, P. (2005). The concept of knowledge work revisited. *Journal of Knowledge Management, 9*(3), 116–127. https://doi.org/10.1108/13673270510602818

Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation–augmentation paradox. *Academy of Management Review, 46*(1), 192–210. https://doi.org/10.5465/amr.2018.0072

Ramírez, Y. W., & Nembhard, D. A. (2004). Measuring knowledge worker productivity: A taxonomy. *Journal of Intellectual Capital, 5*(4), 602–628. https://doi.org/10.1108/14691930410567040

Ruschemeier, H. (2025). Generative AI and data protection. *Cambridge Forum on AI: Law and Governance, 1*, 1–14. https://doi.org/10.1017/cfl.2024.2

Schlee, D., Foerstl, K., & Gutmann, T. (2024). Buying knowledge-intensive business services: Overarching themes, actor relationships, and future research agenda. *European Management Review, 21(4), 760–781.* https://doi.org/10.1111/emre.12691

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research, 104*, 333–339. https://doi.org/10.1016/j.jbusres.2019.07.039

Trincado-Munoz, F. J., Cordasco, C., & Vorley, T. (2025). The dark side of AI in professional services. *The Service Industries Journal, 45*(5–6), 455–474. https://doi.org/10.1080/02642069.2024.2336208

Vaccaro, M., Almaatouq, A., & Malone, T. W. (2024). When combinations of humans and AI are useful: A systematic review and meta-analysis. *Nature Human Behaviour, 8*, 2293–2303. https://doi.org/10.1038/s41562-024-02024-1

von Nordenflycht, A. (2010). What is a professional service firm? Toward a theory and taxonomy of knowledge-intensive firms. *Academy of Management Review, 35*(1), 155–174. https://doi.org/10.5465/amr.35.1.zok155

Werr, A., & Stjernberg, T. (2003). Exploring management consulting firms as knowledge systems. *Organization Studies, 24*(6), 881–908. https://doi.org/10.1177/0170840603024006004

Whittemore, R., & Knafl, K. (2005). The integrative review: Updated methodology. *Journal of Advanced Nursing, 52*(5), 546–553. https://doi.org/10.1111/j.1365-2648.2005.03621.x

Used Artificial Intelligence

Anthropic. (2026). Claude [Large language model]. https://claude.ai/

OpenAI. (2026). ChatGPT [Large language model]. https://chat.openai.com/