UMEÅ SCHOOL OF BUSINESS, ECONOMICS AND STATISTICS
UMEÅ UNIVERSITY

The Invisible Bottom Line

How GitHub Copilot Reshapes Decision-Making and Operational Workflows in Energy Trading Operations

Joel Dalmalm, Eric Jungwallius

Department of Business Administration

International Business Program

Degree Project, 30 Credits, Spring 2026

Supervisor: Catherine Lions

(This page was intentionally left blank)

Abbreviations

AI - Artificial Intelligence

Gen-AI – Generative Artificial Intelligence

ML – Machine Learning

LLM/LLMs – Large Language Model(s)

GPTs – General-Purpose-Technologies

MCP – Model Context Protocol

M365 – Microsoft 365

API – Applications Programming Interface

MIS – Management Information System

RQ – Research Question

Sub-RQ – Sub-Research Question

QA – Quality Assurance

KPI – Key Performance Indicator

IT – Information Technology

CFO – Chief Financial Officer

BA Markets – Business Area Markets

OU – Operations Unit

Abstract

Generative AI tools such as GitHub Copilot are being rapidly embedded into enterprise workflows, yet the literature on their organizational and operational value remains underdeveloped. Existing research focuses primarily on individual developer productivity in controlled settings, while the mechanisms through which AI coding assistants reshape decision-making, workflows, and cost structures in real organizational settings have received limited attention. The financial assessability of these tools has been treated as a presupposition rather than an empirical question.

This study addresses this gap through a qualitative case study of GitHub Copilot's implementation within the Operations function of Vattenfall BA Markets, an energy trading and portfolio optimization environment where data accuracy and process reliability carry direct financial consequences. The study poses one main research question on behavioral and operational implications, and one diagnostic sub-question on the assess ability of financial implications. Empirical material was gathered through semi-structured interviews with ten participants distributed across four blocks corresponding to organizational levels: developers, agile coaches, operations manager, and BA Markets executive management. Data was analyzed thematically and interpreted through a sequential four-theory framework integrating Real Options Theory, Bounded Rationality, Task-Technology Fit, and the Information Systems Success Model.

Five findings emerge. GitHub Copilot has produced substantial individual productivity gains, accompanied by a structural shift in developer work towards upstream specification and downstream verification. These gains are unevenly distributed because the relevant skills were never formally taught. Faster coding has not translated into faster end-to-end delivery; a bottleneck has formed at the quality assurance and code review stage, and adoption travelled below the visibility of the team-coordination layer. Investment and governance decisions have been made through satisficing under low-information conditions and framed in option-theoretic terms. Finally, the financial implications of the implementation are not currently assessable: cost is structurally invisible inside an enterprise contract; benefits are quarantined at the individual level, and downstream quality constraints cap further value capture.

The study contributes a replicable analytical lens for evaluating Generative AI investments in regulated, data-intensive operational environments. It also reframes the financial assessability of such tools as a measurement problem rather than a presupposition.

Keywords: Generative AI; GitHub Copilot; AI-assisted software development; decision-making; task-technology fit; real options theory; energy trading operations.

Acknowledgement

We would like to express our sincere gratitude to Vattenfall BA Markets for their openness, collaboration, and willingness to share their experiences with us.

We are equally grateful to our supervisor, Catherine Lions, for her invaluable guidance and support throughout this work.

Stockholm, 15th of May 2026

Joel Dalmalm & Eric Jungwallius

Table of Contents

(This page was intentionally left blank)

1. Introduction

The first chapter of this study introduces the subject of AI-assisted software development tools and their organizational implications. It provides a background to the emergence of generative AI coding tools and contextualizes the deployment of GitHub Copilot within Vattenfall BA Markets Operations function. The chapter identifies the research gap, presents the problematization and formulates the research question. The purpose of the study is defined, followed by a discussion of delimitations and the study's theoretical contribution.

1.1 Background

Artificial intelligence (AI), a transformative technology that has rapidly emerged across almost all industries, reshaping how organizations operate, make decisions, and create value. In recent years, advancements in machine learning (ML) and Generative AI (Gen-AI), two different branches of AI, have enabled firms to automate tasks, enhance analytical capabilities, and support knowledge-based work at an unprecedented scale. As a result of this, AI adoption has become a strategic point and priority for many organizations seeking to improve efficiency, innovation capacity, and competitiveness in increasingly digitized markets (Babashahi et al., 2024, pp. 1-5).

Despite the growing integration of Gen-AI into organizational processes, evaluating its economic impact has become a significant challenge. Unlike traditional investments in physical assets or financial instruments, Gen-AI does not always produce clearly identifiable outcomes that can be directly observed or measured in financial statements. Many of the benefits associated with Gen-AI adoption such as improved decision-making, enhanced process quality, knowledge augmentation, and operational efficiency are often intangible, not isolated or embedded within existing workflows. Consequently, the value generated by Gen-AI is rarely recognized as a distinct item in the income statement, so it remains somewhat unexplored for organizations and researchers to assess its financial contribution using conventional accounting measures (Corrado et al., 2021, pp. 435-458).

Among these developments, Gen-AI has attracted the authors' attention for its capacity to augment knowledge work by producing code, text, and analytical outputs based on basic language interaction (Dwivedi et al., 2023, pp. 1-5). By embedding it into professional workflows, Gen-AI tools such as GitHub Copilot represent a shift from AI as a standalone analytical system to AI as an integrated co-production resource within daily operational tasks (Dell Acqua et al., 2023, pp. 1-5).

Previous research has highlighted that technologies often create value through process transformation rather than through financial outputs (Kandaurova, 2025, pp. 1-5; Saunila et al., 2023, pp. 1-5). Technology-driven improvements frequently manifest as reduced operational friction, enhanced productivity, improved resource utilization, and improved resource allocation across organizational functions (Vial, 2019, pp. 118-144). However, capturing these effects requires analytical approaches that move beyond the overall financial performance and instead examine how work is performed at the process level. This aligns with the growing research interest in operational value creation, where the focus lies on measuring changes in workflows, task execution, and resource allocation as indicators of technological impact.

Within this context, time efficiency and cost structures have become central metrics for evaluating digital transformation initiatives. By analyzing how technologies influence the duration, complexity, and resource requirement of specific tasks, it becomes possible to estimate productivity gains and cost reductions. Such process-based approaches are particularly relevant for Gen-AI, as many Gen-AI tools are designed to support or partially automate existing workflows rather than replace entire organizational functions.

Given the increasing integration of Gen-AI and the ongoing difficulty of quantifying its financial value, there is a need for empirical studies that examine how the integration of Gen-AI influences decision-making behavior, and not only operational processes and cost structures. Understanding whether, how, and where Gen-AI contributes to measurable economic value creation, and how it reshapes the choices made at individual, team, and managerial levels, is needed to better ground theoretical frameworks in observable organizational reality. As traditional financial metrics prove to be insufficient in capturing the ways Gen-AI shapes cost structures, workflows, and decisions, a deeper interpretation becomes necessary; one that this study shall pursue through a qualitative approach.

1.2 Problematization & Research Gap

Existing literature tends to treat AI adoption and AI value creation as equivalent, assuming that integrating AI technology is sufficient to generate competitive advantage. But this view combines technological investment with realized value, a distinction that has been shown to matter significantly in the context of prior general-purpose technologies (GPTs) (Bresnahan & Trajtenberg, 1995, p. 84). AI, much like earlier digital transformations, has not generated value independently; it is embedded within the organizational processes, human capabilities, and strategic context that mediate its impact (Davenport & Ronanki, 2018, pp. 110-112).

Academic literature has approached AI value creation from several angles. Some researchers have focused on the productivity gains AI produces through automation (Autor, 2015, pp. 5-9), while others have emphasized the strategic value of AI-enabled prediction and decision support (Agrawal et al, 2018, pp. 13-24). Additional papers have examined the role of complementary organizational assets, such as data infrastructure, talent, and management practices, in enabling AI value realization (Brynjolfsson et al., 2019). These angles, while individually insightful, have developed largely independently from one another. As a result, there is little understanding of the mechanisms through which AI creates financial value and where within the organizational value chain this materializes.

Organizations investing in AI tools have little knowledge of how to identify which activities are generating returns, making it difficult to allocate resources rationally, justify continued investment, or divest from underperforming deployments. From a Real Options Theory perspective (Dixit & Pindyck, 1994, pp. 17-34; Myers, 1977, pp. 147-155), an investment can be understood as acquiring a bundle of future options whose value depends on understanding where and how the tool is generating operational impact. Without this understanding, organizations cannot exercise these options effectively, and the strategic flexibility that AI is supposed to provide remains as unused potential rather than actual benefit.

This tension between investment and realized value is supported by recent empirical observations. Toma (2025) explains that while some businesses are prepared to increase their AI investment significantly, with some forecasts that AI could account for up to 7.6% of the IT budgets by 2027, many organizations remain restrictive, emphasizing uncertainty about the returns such an investment will generate. Toma (2025) notes that the restrictiveness is increased by what he describes as, "scarce evidence regarding the effect of AI on company performance" (Toma 2025, pp. 163), initial reports indicate that the technology appears promising but definitive benefits are still unclear. However, any competitive advantage gained from AI integration is at risk of diminishing as market adoption also increases. The observation of Toma shows a fundamental problem, organizations are being asked to commit substantial resources to AI integration and adaptation without a clear understanding of where and how value is created, showing the need for research that goes further than adoption and integration metrics.

A particular example of this dynamic is the rise of AI-assisted coding tools, of which GitHub Copilot has emerged as one of the most adopted across industries. GitHub Copilot, developed by Microsoft and powered by large language models (LLM), is designed to improve software development workflows by generating code suggestions. While productivity increases have been reported at the task level in controlled settings (Ziegler et al., 2022), the translation of these gains into larger operational outcomes within real organizational settings remains poorly understood. Some studies have largely focused on individual developers' productivity (Ziegler et al., 2022), rather than looking at the organizational and process level outcomes from this. Leaving a gap in understanding how such tools influence operations.

The energy market provides a particularly interesting context for examining the operational value of AI-assisted tools. The European energy market has undergone large structural changes in recent years, with natural gas prices rising to an all-time high in 2022 due to a supply shock originating from the Russian invasion of Ukraine which caused a steep decline in Russian gas deliveries (International Energy Agency (IEA), 2022). Since then, volatility has stabilized, but it remains significantly elevated, averaging roughly 50% above the 2010-2019 average in 2024 (IEA, 2025a). Prices have remained elevated through 2025 and into 2026, with renewed price spikes driven by the US-Israeli strikes on Iran which raised concerns over shipment flows through the Strait of Hormuz (IEA, 2025b; McWilliams & Zachmann, 2026). This has fundamentally changed the competitive demands placed on energy trading operations. Commodity markets have now experienced a cycle of high volatility since 2020, driven by multiple disruptions including the COVID-19 pandemic and the European gas crisis, and current trends suggest the elevated volatility could recur in significantly short cycles in the future (McKinsey, 2025). In this market, the speed and accuracy of quantitative analysis, forecasting, and model development have become direct determinants of commercial performance. Companies with the operational capability to process and act on data faster than competitors hold a measurable financial advantage, making the productivity effects of tools like GitHub Copilot not merely an operational question, but a financially consequential one.

This financial perspective of operational performance is underexplored in the AI value creation literature. The Bounded Rationality framework (Simon, 1955, pp. 99-114) discusses a useful perspective here: decision-making in complex, data-intensive environments is constrained by cognitive limits and information processing capacity. AI tools that expand these limits possess the potential to improve the quality and speed of consequential operational decisions. However, to what extent such cognitive enhancement translates into measurable operational and financial outcomes within a specific organization has to the authors' knowledge little to no previous systematic examination.

The energy market is characterized by high capital intensity, regulatory complexity, and increasing volatility driven by the energy transition, geopolitical uncertainty, and the growing share of renewable energy generation. All these different conditions make the sector highly dependent on advanced data processing, forecasting, and decision support capabilities, which has accelerated the adoption of digitalization and AI-based solutions across operations, trading, and risk management. The energy market provides a relevant empirical context for this thesis. Gen-AI has the potential to influence financial planning, operational efficiency, and strategic decisions-making in an environment where timely data-driven insights are critical to value creation and risk mitigation.

Vattenfall is a large, state-owned European energy company and one of the largest in Sweden, with operations spanning power generation, energy trading, distribution, and customer solutions (Vattenfall, n.d.a, n.d.b). Vattenfall is actively investing in digitalization and advanced analytics across its value chain, including initiatives aimed at improving data integration, automation, and analytical capabilities in both operational and corporate functions (Vattenfall, n.d.c; Vattenfall, 2025). Vattenfall is a relevant case for the study on Gen-AI as it represents a capital-intensive, data-rich organization where Gen-AI and related AI applications are being explored to improve operational efficiency, decisions making, and value creation in both the operational and financial contexts (Vattenfall, 2025).

In Vattenfall's BA Markets, operational efficiency, data accuracy, and process reliability have direct financial and strategic consequences. However, how and where a tool like GitHub Copilot creates value in such an environment, and which internal mechanisms explain any observed effects, has not yet been examined in existing research. The literature is strong in describing that AI coding assistance can improve a developer's output, but on the other hand weak on explaining its translation into operational outcomes within specific organizational functions.

The challenges discussed reveal a meaningful gap between the growing organizational integration of AI tools and the understanding of how such tools generate financial value. The uncertainty around AI´s effect on organizational performance is not only a theoretical concern; it also has direct consequences for how organizations invest in and evaluate AI initiatives (Toma, 2025). Therefore, this study aims to address two dimensions of this gap: the mechanisms through which an AI coding assistant reshapes decision-making and operational workflows in a real organizational setting, and the conditions under which the financial implications of such an implementation can be assessed within the organization. GitHub Copilot's implementation within Vattenfall BA Markets' Operations function serves as the empirical point of departure.

1.3 Research Question

The two dimensions mentioned earlier relate to two questions that are often combined in literature, but in this study, the authors have chosen to address them separately. The first question will explore how Gen-AI tools have altered how decisions are made, how tasks are carried out, and how these changes have moved through different teams and management levels within the organization.

The second question deals with whether and how these changes lead to financial results. Though, the authors are aware that the ability to answer this question depends heavily on the organization's measurement systems during the time of the study. This means that the sub-question is not only about the financial outcomes of the integration of Gen-AI, but also about the conditions for assessment within the organization.

Both stated research questions concern effects that arise at different organizational levels that must travel across those levels (individual, team, management, and strategic levels) for it to register as an organizational outcome. Literature often examines such effects at just one level, treating aggregation as transparent. However, the empirical data from this study suggests otherwise: the transitions between different organizational levels is where value is either created or lost. The authors refer to this as the level problem, which is further discussed in Chapter 7. The authors call this the level problem because it influences both research questions: the main question considers effects across multiple levels, and the sub-question treats the conditions for financial assessment as a level-dependent phenomenon.

Main Research Question:

How has the implementation of GitHub Copilot influenced decision-making behavior at Vattenfall BA Markets, and what operational implications does this carry for the organization?

Sub-question:

To what extent can the financial implications of this implementation be assessed within the organization, and what does the answer to that question reveal about how generative AI investments should be evaluated?

The main question is descriptive and exploratory. It aims to map behavioral and operational changes across the organizational levels at which GitHub Copilot is used, and to identify the causes that produce or constrain that change. On the other hand, the sub-question is diagnostic and prospective. Rather than presupposing that financial implications can be quantified from the interview data, it treats their assessability as itself an empirical matter and uses the gap between observed individual-level effects and observable organizational-level outcomes to surface what would be required for a financial assessment to be possible. When framed this way, the inability to settle the financial question with current evidence is not a limitation of the study, but part of its contribution to how AI-tool investments should be evaluated in practice.

1.4 Research Purpose

The purpose of this study is to provide a deeper understanding of how the implementation of GitHub Copilot has influenced operational outcomes within a real organizational context. Adding to this main purpose, this study will also examine a closely related sub-question: to what extent can the financial implications of the implementation be assessed within the organization, and what does the answer to that question reveal about how AI-tool investments should be evaluated moving forward. The sub-question is diagnostic rather than evaluative, which means that it delves into the conditions under which a financial assessment would be possible instead of presupposing that the study's qualitative design is sufficient enough to gather the correct data to conduct a financial assessment.

This study will look beyond deployment metrics to examine what actually changes in processes, workflows, and performance, when an AI coding assistant is introduced into an operationally intensive environment. The energy sector, and energy trading operations in particular, provides an interesting context for such an investigation, given the increase in demand for process/code reliability, and data accuracy (Davenport & Ronanki., 2018, pp. 110-112).

Existing research on AI coding tools has largely been conducted in controlled settings, focusing on individual developer output rather than organizational outcomes (Ziegler et al., 2022). This study will complement and extend that work by examining GitHub Copilot's effect at the functional level, within a specific business area in a large multinational energy company. By doing so, it addresses the recommendations made by Brynjolfsson, E. et al. (2019) for research that goes beyond simple productivity measures and instead investigates how AI creates tangible value within an organization.

This research will theoretically contribute by providing an empirically grounded analysis of AI value creation at the operational level, which from previous research is underdeveloped in current literature. The study will also offer practical contributions by generating suggestions that are directly applicable to organizations facing similar decisions and questions about AI integration in complex and regulated environments. The findings of this study are intended to generate knowledge that reaches beyond Vattenfall itself, in which it seeks to offer insights relevant for other organizations facing similar AI implementation challenges.

1.5 Delimitations

This study is delimited to GitHub Copilot interviews for excluding Microsoft 365 Copilot and other Gen-AI tools present in the organization, and to the operations department within Vattenfall BA Markets. Even though preliminary fieldwork was done and provided access to a broader part of the organization, the empirical focus was narrowed down to the trading and management departments within operations exhibiting the most consistent and substantial adoption of GitHub Copilot.

Methodologically, the study is delimited to a qualitative case study design based on semi-structured interviews. Quantitative productivity benchmarking and telemetry-based measurement of code output were deliberately not pursued, as the research questions targets how and where financial and operational value is created rather than how much. The analysis is further restricted to behavioral and operational mechanisms at the organizational level. The technical quality of the code produced, and individual-level psychological constructs fall outside the scope.

The financial dimension is treated diagnostically rather than evaluatively. The study does not attempt to quantify a return on investment but instead examines the conditions under which such an assessment would be possible. The theoretical framing is likewise delimited to four chosen frameworks applied as a sequential analytical chain, while other AI-adoption frameworks are acknowledged but not incorporated.

Finally, the participants' scope is limited to active users and managers of GitHub Copilot within Vattenfall BA Markets and does not include perspectives from group-level finance, IT security, risk management, or the vendor. The findings are situated within the regulated European energy-trading context.

1.6 Theoretical Contribution

This study makes a theoretical contribution by applying and connecting four established frameworks, Real Options Theory, Bounded Rationality, Task Technology Fit, and the Information System Success Model, to the expanding market of AI-assisted coding tools in operational and financial environments. While each of these four theories has been widely applied in management research, they have rarely been used in combination. Few, if any, prior studies that have integrated them into a full analytical lens for evaluating how an AI coding assistant generates value within a financially intensive operational environment. By doing this, the study advances the understanding of how AI implementation moves beyond deployment metrics, towards measurable operational outcomes, addressing a gap identified by Brynjolfsson et al. (2019) in research on AI and organizational performance.

In addition to applying these four frameworks to the case of an AI coding assistant in a financially intensive operational environment, the study contributes to a methodological argument about the conditions under which the financial implications of Gen-AI tools can be assessed. This argument, developed in Chapter 7 as the level problem, addresses the sub-research question directly and responds to a call in the literature for research that goes beyond simple productivity measures and investigates the organizational mechanisms by which AI creates tangible financial value (Brynjolfsson et al., 2019).

1.7 Practical Contribution

Practically, this study will provide insights and a study instrument that can be directly applicable to organizations evaluating AI tools for investments in operationally intensive settings. This study will seek to offer an insight into the reasoning behind an AI tool investment that moves beyond developer productivity. For financial decision makers, the findings will provide a richer picture of where AI coding tools create value. This is relevant as previous literature within the subject by Toma (2025) notes that organizations are willing to commit resources to AI without clear frameworks for evaluating where value materializes. Beyond the findings, this study will contribute a replicable diagnostic approach. The multi-level, block-based interview design is itself a transferable instrument that other firms can apply to locate where an AI tool creates or loses value across their organizational levels. The findings, together with the approach used to produce them, are intended to go beyond Vattenfall to CFOs and operations managers in other regulated, data-driven environments facing similar AI adoption decisions.

2. Conceptual Framework

The second chapter sets out the conceptual foundations of the study, explaining generative AI and GitHub Copilot, the empirical context of Vattenfall BA Markets, and how AI-assisted tools shape decision-making in technical workflows along with the operational and financial implications that may follow.

2.1 What is Generative AI?

Generative Artificial Intelligence, or more commonly referred to as Gen-AI, represents a subfield of artificial intelligence, characterized by its capabilities to produce original content rather than merely classifying or predicting within predefined parameters. As earlier AI technologies were designed to automate defined rule-based tasks, Gen-AI uses generative models to produce texts, images, videos, and software code by learning underlying patterns from large datasets and generating new data in response to natural language prompts (Amazon Web Services, 2024). For most Gen-AI tools to function, they need Large Language Models (LLMs) which are deep learning models trained on vast amounts of data and are capable of understanding and generating natural language across a wide range of tasks (IBM, 2023).

The emergence of Gen-AI has marked a new era in natural language processing, with its current capabilities now considered crucial for researchers, practitioners, and policymakers shaping the responsible integration of these technologies into organizational workflows (Hagos et al., 2024, pp. 1-5). It is within this context that tools developed by Microsoft, such as Microsoft Copilot M365 and GitHub Copilot have emerged, applying Gen-AI capabilities directly to enterprise workflows and forming the subject of inquiry in this thesis.

2.1.1 What is Microsoft Copilot and GitHub Copilot?

Microsoft Copilot represents a range of AI-assisted tools developed through Microsoft's partnership with OpenAI, with a main design aimed to embed generative AI capabilities into enterprise workflows. Microsoft 365 (M365) Copilot is a sophisticated processing engine that combines the power of large language models (LLMs) with the M365 application suite and an organization's business data stored in the Microsoft Graph (Spataro, 2023). The main factor that differentiates it from general-purpose AI assistants is that the model is grounded in the specific organizational context by drawing in information from real-time content such as the user's emails, documents, calendar, chats, and meetings, delivering responses relevant to the specific application being used (Microsoft, 2024).

M365 Copilot is integrated across the core applications related to the Office 365 package that organizations such as Vattenfall use daily. The applications include Word, Excel, PowerPoint, Outlook, and Teams, where Copilot assists users in creating documents, suggesting formulas, summarizing texts and meetings.

GitHub Copilot, while sharing the same technological origin, takes on a different function which is oriented mainly towards software development. Its model has been trained and developed on natural language text and source code from publicly available sources, including code in public repositories on GitHub (GitHub, 2025). The tools' primary function is to act as an AI programmer by either complementing code as it is written or generating code from the users prompts (GitHub, 2025). According to GitHub themselves, developers using Copilot are up to 55% more productive at writing code without risking a drop in overall quality (Peng et al., 2023, p. 1).

Confusion may arise when discussing Copilot as both tools fall under the broader brand, but their role in organizations differs considerably. M365 Copilot is used primarily by individuals engaged in analysis, communication, and reporting, which are functions that are central to certain departments within BA Markets. On the other hand, GitHub Copilot is mainly used by technical staff working within software development and quantitative analysis as these individuals build and maintain quantitative models, automation scripts, and data pipelines. Functions which are central for the operations and trading departments.

2.1.2 The Rise of AI Augmentation in Financial and Operational Workflows

The adoption of Generative Artificial Intelligence (Gen-AI) and Machine Learning (ML) into financial and operational workflows points towards one of the most significant technology shifts of the current decade. In 2025, a proprietary McKinsey survey sent to 102 CFOs showed that 44 percent of respondents said that they used Gen-AI for over five use cases in 2025 and 65 percent of respondents mentioned their organizations will increase their Gen AI investments in 2025; in 2023, only a about a quarter of respondents said the same (Sukharevsky et al., 2025). As the field expands and advancements in computing and algorithms broaden the use cases for these technologies, AI has reshaped various sectors, particularly finance. The growing adoption of AI tools in financial markets is largely attributed to their ability to enhance operational efficiency and foster innovation, with primary applications encompassing algorithmic trading, risk mitigation, and credit scoring (Jiang et al., 2023, p.434). This trend is not confined to traditional financial institutions but has reached energy companies operating in commodity markets with similar needs to optimize trading strategies, improve forecasting accuracy, and manage the inherent complexity of volatile markets (Ghoddusi et al., 2019, pp. 709-712).

Within energy and commodity markets specifically, AI augmentation has increasingly taken hold through quantitative and algorithmic applications. Algorithmic trading is becoming the main approach in many market-based firms and is more commonly known to be used in financial institutions such as hedge funds. However, the adoption of AI and ML tools has gone well beyond traditional financial actors and landed in the energy sector more broadly. A comprehensive review of ML approaches in energy trading identifies a wide range of use cases where AI applications can be used, in which reinforcement learning gained the most traction as a method for optimizing decisions in dynamic and uncertain market environments (Khalid et al., 2024, pp. 16-17). The scale of the current transformation is significant, with the global AI in energy market valued at USD 18.1 billion in 2025 and projected to reach USD 75.53 billion by 2034, growing at a compound annual growth rate of 17.2 percent (Precedence Research, 2025).

At the operational level, the augmentation of human workflows through AI has become a defining characteristic of modern organizations across sectors. McKinsey estimates the long-term AI opportunity at USD 4.4 trillion in added productivity growth potential from corporate use cases, which highlights the reach that the technology has beyond any single industry (Mayer et al., 2025). As more research and studies involving the AI transformation within corporations' increase, so does the measurability of the transformation. Companies with well-developed AI-based initiatives have been found to grow their revenues 2.5 times faster and their productivity 2.4 times higher than those without AI-integrated processes, which shows the competitive advantage of integrating such tools (Negrea, 2025). The current implementation of Gen-AI tools across enterprises reveals a preference for augmenting human workflows over full automation, with organizations focusing on embedding AI primarily in software development functions, alongside other operational tasks (Tully, 2024). For organizations in the energy sector, such as Vattenfall (more specifically Vattenfall BA Markets) where their most profitable operational workflows depend heavily on the speed and accuracy of quantitative analysis, forecasting, and model development, this broader trend creates both an imperative and an opportunity. The competitive edge increasingly depends not only on market knowledge but on the quality of the tools that support it.

2.2 Vattenfall BA Markets

Vattenfall BA Markets is the energy trading and portfolio optimization arm of Vattenfall, one of Europe's largest energy companies, with approximately 21,000 employees in total. BA Markets maximizes the value of Vattenfall's generation and customer portfolios by optimizing, hedging, sourcing, and trading electricity, fuels, emissions, freight, and renewable certificates. It serves as Vattenfall's single point of access to European energy commodity markets and employs roughly 500-560 professionals across Sweden, Germany, and the Netherlands (Vattenfall, 2026).

Within BA Markets, the Operations Unit (OU Operations) plays a central role in enabling digital transformation. OU Operations consists of skilled technical teams which are responsible for developing and maintaining data pipelines, data, price modelling, IT applications, and frameworks that assist in trading, forecasting, and risk management processes. Within this context, if model errors, data delays, or system failures occur, the organization can find themselves in a suboptimal position or an increased risk exposure, which makes this environment particularly interesting for evaluating Gen-AI tools such as GitHub Copilot (Vattenfall, 2026).

2.3 Decision-Making in Technical and Operational Workflows

In the organizational setting, decision-making is rarely confined to discreet, high-level choices made by senior management. In both technical and operationally intensive business functions, it also encompasses the continuous stream of task-level judgements made by analysts and developers throughout their daily work. Micro-decisions, made repeatedly across a team or division, collectively shape the quality and reliability of outputs that ultimately carry both financial and operational consequences.

Research shows that as cognitive demands accumulate, decision quality deteriorates. Individuals rely more heavily on heuristics, are more susceptible to bias, and are less able to engage in deliberate analysis (Kahneman, 2011, pp. 29-49). Due to this, the study becomes specifically relevant within energy trading, as analysts must maintain accuracy across the decisions that are made and routinely process large volumes of data and operate under constant pressure. The risk of confirmation bias (the tendency to favor information that aligns with pre-existing beliefs) is well-documented and has consequences for the decisions that are taken (Fonseca Costa et al., 2020; Kappes et al., 2020, p. 130).

In trading-oriented environments, errors can spread through algorithms and valuation pipelines, which create distortions that affect risk assessments and bottom-line performance. Previous literature within behavioral finance shows that decision quality under uncertainty is vulnerable to cognitive strain. Cognitive biases such as overconfidence, information overload, and anchoring have the possibility to shape how decision makers interpret information, which might lead to negative outcomes (Barberis & Thaler, 2003, pp. 1053-1055, 1063-1069; Tversky & Kahneman, 1974, pp. 29-49).

The usage of decision-support systems will not be able to accelerate productivity universally. However, they can assist in shaping how problems are framed and how alternatives are evaluated. Research within automation bias and human-AI collaboration highlights that individuals under time pressure or increased workload are more prone to place a higher trust in model-generated outputs (Mosier & Skitka, 1996, pp. 201-220; Parasuraman & Riley, 1997, pp. 230-253). Within the context of operations, this dynamic raises an important question about how AI-generated suggestions may influence judgement, to the extent to which cognitive load is redistributed or reduced, and how such shifts affect financial accuracy and operational reliability.

2.4 Operational Value Creation and Its Financial Implications

For AI tools embedded in knowledge-intensive workflows, value creation rarely appears as a discrete financial event. Instead, it accumulates at the process level: tasks can be completed more efficiently, errors reduced, output quality improved, and skilled workers freed from routine work to focus on higher-level activities. These challenges are real and economically meaningful, but they require process-level observation to detect and are not easily captured in conventional financial reporting (Corrado et al., 2021, pp. 435-458).

This distinction between operational value and financial value is important for framing the study. Operational value refers to measurable changes in how work is performed: this could be time saved, error rates reduced, or throughput increased. Financial value refers to outcomes that appear in the income statements or balance sheets. The relationship between the two is real but indirect. Operational improvements need to accumulate to a sufficient scale, and in the right areas, before they register as financial outcomes. The value that can be squeezed from the implementation of AI tools is dependent on where and how it is deployed within specific organizational workflows, not simply adoption (Mayer et al., 2025).

For Vattenfall BA Markets, the implication is clear. GitHub Copilot does not generate revenue directly, nor does it appear in any profit and loss statement. But if GitHub Copilot accelerates and optimizes model development, or reduces the time analysts spend on routine tasks, it reduces the cost of producing operational outputs and potentially improves their quality, both of which carry financial consequences for an organization operating in volatile energy markets. Understanding whether and how this chain of value creation operates within the operations function is the focus of the main research question. Whether the financial end of the chain can be assessed at all under current organizational conditions is the focus of the sub-question.

3. Theoretical Framework

This chapter will discuss the theoretical framework that has been chosen for the study. The purpose of the framework is not to summarize the existing theory, but to construct an analytical lens in which the empirical findings can be interpreted in relation to the stated main and sub research questions. As these questions have different layers, each theory has been selected because they contribute to the overarching theme and question. The purpose of the theoretical framework is to use existing theories to understand the logic of the investment, the cognitive and organizational mechanisms it triggers, and the conditions under which these mechanisms produce measurable outcomes. The chapter concludes with a conceptual summary that illustrates how the theories relate to one another and to the research questions:

How has the implementation of GitHub Copilot influenced decision-making behavior at Vattenfall BA Markets, and what operational implications does this carry for the organization?

To what extent can the financial implications of this implementation be assessed within the organization, and what does the answer to that question reveal about how generative AI investments should be evaluated?

3.1 Real Options Theory

Real Options Theory (ROT) is a framework which originates in corporate finance for evaluating investment decisions made under conditions of uncertainty. The foundation of the theory, introduced by Myers (1977, pp. 147-155), is that many corporate assets, in particular growth opportunities, can be understood as call options on the firm's future investments. In the same way a financial option gives you the right but not the obligation to buy an asset at a predetermined price, a real option gives the firm the right but not the obligation to pursue a future investment if the laid conditions seem favorable (Myers, 1977, pp. 147-155). The implication of this is that the value of an investment cannot be evaluated solely by its immediate and perceived returns, as it may also generate future strategic avenues which are contingent on how circumstances evolve.

Traditional net present value (NPV) analysis treats investment decisions as static and irreversible whilst ROT explicitly recognizes flexibility as a source of economic value. Dixit & Pindyck (1994, pp. 17-34) delved deeper into this by showcasing that uncertainty can increase the value of a strategic option rather than the uncertainty decreasing it. This is due to the fact that when a decision can be deferred or abandoned, greater uncertainty will ultimately increase the value of acting at the optimal moment, while on the other hand, the downside remains bounded to the choice of not exercising the option at all (Dixit & Pindyck, 1994, pp. 17-34). Other studies have used ROT well beyond its origins in natural resource economics and capital budgeting, in which is has also been to strategic decisions in technology-intensive environments (Li et al., 2025, pp. 631-640).

In this study, ROT helps explain the reasoning behind the GitHub Copilot investment within Vattenfall BA Markets. As the investment cannot be treated as a one-time expenditure that is evaluated against basic productivity gains, ROT will direct attention towards the strategic option that the investment creates: Vattenfall BA Markets commits time and resources today for the possibility of further strategic choices in the future. An example of this would be to expand their existing AI infrastructure, building models internally that competitors cannot easily replicate or adapt to other technological implementations from a position of prior experience. Li et al. (2025) notes that organizations which apply real options thinking into their decision-making processes are better positioned to mitigate downside risks while retaining the possibility to capitalize on future opportunities (Li et al., 2025, pp. 634-635). ROT may indicate why the operational and financial implications of the GitHub Copilot implementation may not be fully visible in short-term metrics, and which is why a broader analytical lens is appropriate for this study.

3.2 Bounded Rationality

Bounded Rationality was introduced by Herbert Simon (1955) as a concept used in both economics and organizational theory. Simon questioned the assumption that individuals make decisions by processing all available information rationally and selecting the option that maximizes their utility (Simon, 1955, p. 99). This framework developed by Simon has since become a cornerstone to behavioral economics and behavioral finance, where it underpins challenges to the rational-actor and efficient-market assumptions. The main arguments form around the fact that decision-makers work within three core constraints: limited information, cognitive processing capacity, and time. Due to this, individuals search for a solution that is "good enough" given the constraints they are under, rather than, in theory, finding the best solution possible (Simon, 1955, pp. 99-114). This distinction can assist in describing how organizations function, since the quality of decisions made at every level is bound by the cognitive limitations of the individuals making them.

Kahneman & Tversky (1979) identified specific cognitive biases as expressions of bounded rationality, depicting systematic and predictable patterns through which human judgment deviates from the rationale norm (Kahneman & Tversky, 1979, pp. 263-284). Examples of this consists of overconfidence in one's own estimates, anchoring to initial information, and the tendency to weight losses more heavily than gains (Kahneman & Tversky, 1979, pp. 270-280). In software development, where complex technical problems occur, these stated biases could have a higher chance of manifesting. This could lead to suboptimal decisions, underestimations of various task complexities, and over-reliance of familiar solutions that might not be best suited for the problem at hand.

GitHub Copilot should function as a cognitive aid that can partially offset the limitations Simon identified. As the Gen-AI can produce relevant code suggestions instantly, lower the overall cognitive load associated with coding tasks, and provide a form of real-time external reference, the tool should be able to counteract various individual biases and knowledge gaps. If bounded rationality correctly describes the constraints under which developers at Vattenfall BA Markets operate, then a tool that alleviates these constraints should be able to produce observable changes in how they approach problems. Bounded rationality also provides a bridge between individual-level behavior and organization-level outcomes. If a significant number of individuals make better decisions more efficiently, the overall payoff effect should both show an operational and financial improvement for the organization.

3.3 Task-Technology Fit

According to Goodhue & Thompson (1995), who developed the Task-Technology Fit Theory (TTF), a technology will improve individual performance only to the extent that it fits the tasks the user needs to perform (Goodhue & Thompson, 1995, p. 213). ”Fit” in this framework is the degree to which a technology’s functionality matches the requirements of the tasks it is applied to. When ”fit” is high, the technology augments the user’s capability in ways that are directly relevant to their work. In contrast, when the ”fit” is low, the technology may add cognitive burden, require workarounds, or simply go unused regardless of how sophisticated or well-designed the tool is in isolation (Goodhue & Thompson, 1995, pp. 213-216). TTF therefore shifts the analytical focus from technology characteristics alone to the interaction between technology, tasks, and users.

This theoretical lens is particularly important for the scope of this study which focuses on GitHub Copilot’s implementation across a diverse organization. As the tool is used primarily by developers at Vattenfall BA Markets, not everyone will perform the exact same task. Some may work primarily on routine, well-defined coding problems where Copilot’s autocomplete and suggestion capabilities align closely with their needs, while others may engage in highly complex, novel, or domain-specific work where the tool’s suggestions are less reliable or relevant. The theory itself predicts that the behavioral and operational effects of GitHub Copilot will vary accordingly, and this variation is empirically meaningful rather than noise to be averaged away. It should be noted that TTF is directly relevant only to the interview blocks involving developers who use GitHub Copilot in their own tasks, and not to the management and strategic blocks where the tool is not used hands-on. The authors include TTF because the user-task level is where the tool’s effects originate, and any account of how they propagate upward must begin there. The partial applicability of TTF across the interview structure is therefore a deliberate feature of the framework rather than a limitation of it.

It is important to note that TTF was originally developed in the context of traditional IT-supported decision-making tasks, and its application to an AI-assisted coding tool represents an extension of the framework beyond its original empirical base (Goodhue & Thompson, 1995, p. 213). Furthermore, TTF complements the Bounded Rationality perspective and ROT introduced earlier. When the “fit” between GitHub Copilot and a developer’s task is low, instead of helping the developer think more clearly, the tool can make things worse and may reinforce rather than reduce the biases. From a ROT perspective, if the tool does not “fit” the tasks developers are doing, the future strategic value and flexibility for the organization never materializes, leading to the “option” existing on paper while it cannot be exercised in practice.

3.4 The Information Systems Success Model

The Information Systems Success Model, proposed by DeLone & McLean (1992) and subsequently updated by DeLone & McLean (2003), proposes a multi-dimensional framework for evaluating the success of an information system in an organizational context. The original model identified six main dimensions of success – system quality, information quality, use, user satisfaction, individual impact, and organizational impact - which DeLone & McLean (2003) later restructured by adding service quality and consolidating individual and organizational impacts into a single dimension, net benefits (DeLone & McLean, 2003, p. 10). These dimensions are not independent of one another as they form a causal chain in which the quality dimensions shape use and satisfaction, which in turn produce net benefits at both the individual and organizational level (DeLone & McLean, 2003, p. 11). The concept of net benefits is deliberately broad, which has been noted as a limitation of the model, as it leaves room for varied implementation across studies (DeLone & McLean, 2003, p. 24). However, in the context of this study, the breadth of net benefits becomes an asset, as it can encompass decision quality, operational efficiency, and cost reduction – outcomes that are directly relevant to the research question.

It should be acknowledged that the IS Success Model was developed primarily for traditional organizational information systems, and being applied to an AI-assisted coding tool such as GitHub Copilot represents an extension of the original framework. Nevertheless, the model's dimension-based structure remains applicable, as GitHub Copilot can be evaluated across each mentioned dimension. Therefore, the IS Success Model serves as the primary framework for connecting the behavioral changes identified through the earlier mentioned theories to the organizational outcomes specified in the research question. Through the lens of this framework, GitHub Copilot can be evaluated across the aforementioned dimensions: its system quality concerns the reliability, speed, and accuracy of its code suggestions; its information quality concerns the relevance and correctness of the content it produces; its use and user satisfaction are empirically observable through interviews; and its net benefits – the dimension most directly relevant to the research question – encompass the financial and operational implications that the study seeks to examine (DeLone & McLean, 2003, p. 24).

3.5 Theory Discussion

The theories that have been chosen for the study address different layers of the research question, from investing in the technology, through individual cognitive effects, to the conditions under which organizational outcomes emerge. This forms a chain that will help the analysis to interpret the results in the light of the Main and Sub research questions. While each theory contributes to a distinct perspective and formulates a framework relevant to the study, their limitations and suitability deserve a critical reflection.

ROT was originally developed for capital-intensive investments where option values can be formally quantified. Although in this study, the "options" created by the investment into GitHub Copilot are qualitative in nature and cannot be priced in the way the theory's mathematical foundations presuppose. That is why ROT is applied as a concept rather than a valuation tool, an application previously used by Li et al. (2025), but it represents a meaningful departure from the theory's origins.

Bounded Rationality is expected to assist in explaining why developers benefit from cognitive assistance. However, it does not account for the possibility that Gen-AI tools may introduce new cognitive constraints alongside the ones that they relieve. An example could be that the individual creates an over-reliance on the generated suggestions, and/or reduces the incentive to develop a deep technical understanding of the tasks/tools that are carried out/used. The theory needs therefore to be applied with this ambiguity in mind. Nevertheless, it provides a foundation for understanding whether GitHub Copilot expands the decision-making capacity of developers at Vattenfall BA Markets, or whether its benefits are offset by new cognitive dependencies introduced by the tool itself.

TTF (Task-Technology Fit) was first developed for traditional IT decision-support systems and has since not been empirically validated for Gen-AI tools, which differ substantially from rule-based systems in both capability and unpredictability. Its application for this study is justified as the core logic states that a technology value depends on how well it matches the demands of the tasks it is applied to, regardless of whether the tool is rule-based or generative. Despite that, it needs to be understood that the usage of the theory goes beyond the framework's original empirical base.

The authors understand that the usage of four theories simultaneously poses a risk that the analysis becomes diffuse. However, this is managed by treating the theories as a sequential chain rather than independent lenses, ensuring each theory contributes to a distinct and traceable role in addressing the research questions. The following image shows how the four theories are organized as a chain across at which each locates the exploratory work, and the transitions that connect one level to the next. Each theory accounts for a specific mechanism: strategic commitment, individual cognition, task augmentation, and organizational aggregation. The transitions between them represent a conceptual hand-off at which value can be preserved or lost. No single theory in the theoretical framework can see beyond the level at which it operates, which is why the chain is constructed rather than the theories in parallel with each other. It is how the framework is built up, and not the theories taken individually, that makes the transitions between levels analytically visible. The financial aspect threads through this chain at specific points. ROT entering from corporate finance, Bounded Rationality from the tradition foundational to behavioral finance, and the IS Success Model's net-benefits dimension as the point at which operational outcomes would have to aggregate to register financially.

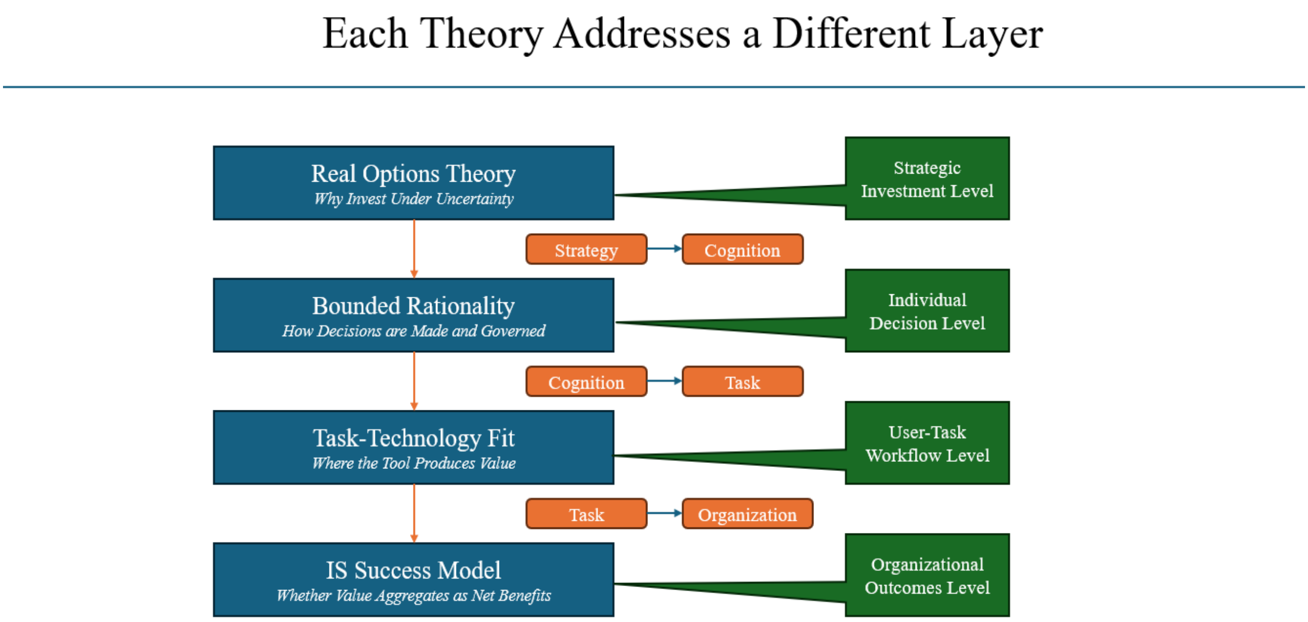

Figure 1. The four theories form a chain rather than independent lenses, each addressing a distinct organizational level; strategic investment, Individual decision-making, user-task workflow, and organizational outcomes. Source: Authors' own construction.

4.0 Methodology

The Methodology chapter outlines and justifies the methodological choices made throughout this study. It begins with the research philosophy adopted by the authors, followed by separate sections addressing the ontological, epistemological, and axiological positions of the study. The chapter then turns to the research approach, research method, and research design before specifying how an event study is integrated within the case study framework. It closes with a statement on the use of AI during the research process and a discussion of the literature base and source criticism applied throughout.

4.1 Research Philosophy

To choose an appropriate research method and have the right point of view, it is fundamentally important to evaluate research philosophies, as they shape the assumptions underlying the entire research process, from how the research question is understood, to how the data is collected, and how the data gathered is interpreted. Saunders et al. (2009, p. 108) gives the description of research philosophy as the system of beliefs and assumptions about the development of knowledge. That it is important to use appropriate research philosophy because of the assumptions that will shape how questions are understood, how the authors use methods and interpret findings. Establishing a clear philosophical foundation therefore ensures that ontological and epistemological assumptions of the study stay constant throughout the entirety of the study. Giving coherence to the research as a whole (Saunders et al., 2009)

This study adopts pragmatist research philosophy. The authors believe that a Pragmatism approach is appropriate for this study since both of the research questions have different demands for evidence. Looking at the main research question, it asks employees how their experience and interpretation of GitHub Copilot and what kind of operational implications that have followed from the introduction. The question has two parts in it, the first require the authors to engage with the emotional and subjective side, and the second part of the question requires more of observing workflows and process changes and seeing these as meaningful evidence. Moving on to the sub research question, it ask whether the financial implication of the implementation can currently be assessed within the organization, this part compared to the other question requires looking at the measurement conditions themselves as empirical material instead of assuming that one piece of evidence will fix the issue. Rather than committing to a single philosophical extreme, pragmatism allows the study to draw on whichever form of evidence best illustrates the phenomenon, provided it contributes to answering the research question (Saunders et al., 2009, p. 109). Ontologically, this means treating reality as partly objective and partly socially constructed; epistemologically, it means judging knowledge by its usefulness in addressing real organizational problems; and axiologically, it means acknowledging the authors proximity to the field and adopting a reflexive stance throughout the analysis. These three dimensions are unpacked in the following sub-sections.

4.2 Ontology

The research ontology is linked to the assumptions about nature and reality (Saunders et al., 2009, p. 134). In social research, reality is not a singular or fixed thing; it can be understood in different ways depending on the perspective of the researcher (Bell, 2022. p. 26). Within ontology, Saunders et al. (2009, pp. 136-138) argues further that there are two angles within ontology, objectivism, and subjectivism. Objectivism is that social phenomena exist independently of human perception and social factors, that external reality exists whether the researcher is present to observe it (Bell, 2022, p. 26). Subjectivism, on the other hand, says that social reality is constructed and seen through the interactions and experiences of the people involved. With this view, the researcher cannot fully let go of their own values and perceptions, as these shape the understanding of the phenomenon being studied (Saunders et al., 2009, pp. 137-138). The main research question in this study asks both how decision-making has changed since the implementation of GitHub Copilot, which will require engaging with the subjective interpretations of organizational members, and the operational effects which followed from those changes. Whilst the sub-question adds a further ontological dimension by questioning whether the financial implications of the GitHub Copilot implementation can currently be assessed, treating the conditions for those measurements as part of organizational reality. This dual focus calls for a pragmatist stance from an ontological point of view, which sees what works best for answering the research questions rather than solely focusing on one philosophical side (Saunders et al., 2009, p. 109). Under pragmatism, reality is viewed as partly objective and partly socially construed. The researcher's task is to engage with both perspectives in order to develop the most meaningful and valuable understanding of the phenomenon.

4.3 Epistemology

Epistemology is the theory that addresses the question of what forms valid knowledge and how that knowledge can be obtained (Bell, 2022, p. 29: Collis & Hussey, 2021, p. 43). Collis argues about two primary epistemological viewpoints in social research: interpretivism and positivism, Saunders et al. (2009, p. 148) also identified realism as a third viewpoint. Positivism is explained as the reality is independent of the researcher and that the goal of the study is objective: to achieve generalization through empirical testing (Collis & Hussey, 2021, p. 40). The positivism in general leads to quantitative methods and hypothesis testing as the tools (Saunders et al., 2009, pp. 145-147). Interpretivism, in contrast, is better explained in the belief that social reality is subjective and shaped by our human perception and experience (Collis & Hussey, 2021, pp. 41- 43). It is to say that the researcher cannot be fully independent of what is being researched, and that the goal is to describe, translate, and interpret the meaning rather than to produce generalizable laws (Collis & Hussey, 2021, p. 41).

In this study, neither a purely positivist nor a purely interpretivist stance is deemed fully appropriate by the authors. The main research question aims to explore how employees of the organization experience and interpret the implementation of GitHub Copilot while also acknowledging that practical implications (such as workflow changes or operational impacts) are meaningful only if they help answer that question. The sub-question, however, raises a different epistemological concern by asking what is currently known about the financial implications of the implementation given the organization's measurement infrastructure, and what the limits of that knowledge reveal about how Gen-AI investments should be evaluated. For these reasons, the study will adopt a pragmatic epistemology stance. Pragmatism holds that the value of knowledge lies in its usefulness and its ability to address real-world problems. This allows different forms of qualitative evidence to be combined when they help illustrate the phenomenon under study (Saunders et al., 2009, p. 109).

Although some supplementary organizational data (KPIs, operational indicators) may be referenced if provided by interviewees, these are used only as contextual material and not as part of a mix-method or quantitative research design. The study remains fully qualitative, relying primarily on semi-structured interviews to understand how individuals make sense of changes in their work, decision-making behavior, and perceived implications. From a pragmatic standpoint, such qualitative accounts offer meaningful insights into how GitHub Copilot is experienced, interpreted, and embedded within Vattenfall BA Markets operational context.

Having a pragmatic stance requires transparency regarding the role of researchers. Interpretation is easily influenced by prior knowledge; therefore, reflexivity is important throughout the analysis. Any supplementary quantitative material mentioned by participants is treated as a supportive context rather than objective measurement of organizational reality. Correspondingly, when interpreting the qualitative data, the authors remain aware of how their own backgrounds may influence the analytical process.

By grounding the study in pragmatism, the research design prioritizes usefulness, contextual understanding, and practical meaning, ensuring that the conclusions resonate with both academic expectations and the organizational realities of Vattenfall BA Markets.

4.4 Axiology

Axiology describes the role of values in research, specifically how the researcher’s own values influence the collection of information, interpretation, and reporting of information (Saunders et al., 2009, p. 116). This is particularly relevant in qualitative studies where detachment from the subject of inquiry is neither achievable nor desirable (Collis & Hussey, 2021, p. 57).

This study will be conducted from a subjective perspective. As the research takes place within an organizational context the authors are familiar with, there is an inherent risk that the authors’ prior knowledge and professional proximity to the subject may influence how interview responses are interpreted. The authors acknowledge this and have sought to minimize it by being transparent in the analytical process, grounding interpretations consistently in empirical data, and subjecting findings to critical reflections against the theoretical frameworks rather than personal judgement.

At the same time, familiarity with the context is not solely a limitation. Familiarity with the operational environment of Vattenfall BA Markets enables more targeted and informed interviewing, which contributes to the depth and relevance of the data collected (Saunders et al., 2009, p. 151). The authors aim for what Hammersley (2012, p. 71) described as a reflexive stance, recognizing the influence of one’s own values while remaining committed to rigorous and transparent inquiry.

4.5 Research Approach

For the authors to be able to collect data that is of high quality and accurate to the thesis, they need to determine an appropriate research approach that also aligns with the purpose of the study (Saunders et al., 2009, p. 154). The two best-known research approaches are inductive and deductive, as both have a big difference in their relationship between theory and empirical data, it is very important to choose one that is suitable to answer the research question.

A deductive approach starts with establishing theory and uses it to form a hypothesis that is tested against empirical observations and data to confirm or reject the theory (Collis & Hussey, 2021, pp. 6-7). In the deductive approach the researchers go from the general to the more specific, collecting data on variables that the theory has already identified as significant, and then evaluating whether the relationship between they hold in practice (Saunders et al., 2009, pp. 156-157). The deductive approach is most appropriate when the research field is developed enough to generate a testable proposition. An inductive approach works in the totally opposite direction. Theory is developed from observations rather than before observations (Collis & Hussey, 2021, p. 8). The researchers approach the field with open questions and allow patterns to form and emerge from the collected data and then build theoretical understanding from there (Saunders et al., 2009, p. 157). Woicesshyn, & Daellenbach, (2018, p. 186) argued that the inductive approach is practically well suited to understanding and capturing how people think and respond to their environment, this is directly relevant when studying the human reaction and organizational outcome of organizational change.

This study primarily adopts an inductive approach for two reasons. Firstly, as established before, peer-reviewed research on the organizational and operational effects of GitHub Copilot in real business context is largely nonexistent (Ziegler et al., 2022). By not having a well-developed body of theory specific to this phenomenon means that a deductive approach (which needs preexisting theory to be able to generate testable hypotheses) would be too early to adopt and not suited the research questions. Secondly, the research question itself points towards where and how value is created, and whether the financial implications of the value can currently be assessed within Vattenfall BA Markets. Both calls for an open exploration rather than the confirmation of pre-specified relationships. The primary data source for this study will be semi-structured interviews with employees working within Vattenfall BA Markets, and especially in the operations department. This is to create and to be able to capture variety as there are many different workflows and people within the operations department. This will give the authors a large variety of answers and experiences, that will allow for different patterns to emerge in the analysis rather than imposing predetermined frameworks upon it.

These patterns will surface through thematic analysis of the interview material, following the six-step procedure developed by Braun & Clarke (2006, pp. 86-93). Interview transcripts will be coded openly, related codes grouped into broader themes, and those themes reviewed and refined against the full dataset before being connected back to the research questions. The full procedures are detailed in section 5.4.

4.6 Research Method

In academic research, the choice of method is determined by the nature of the research question and the philosophical and design choices that precede it (Queiros et al., 2017, p.

369). The two research methods are quantitative and qualitative; each carries potential and limitations that make them more or less appropriate to the study depending on what the researcher is trying to achieve.

Quantitative research focuses on objectivity and the measurement of variables across a very large sample number (Queiros et al., 2017, p. 370). The strength with the quantitative approach is the ability it possesses to identify statistical patterns and relationships, and to produce findings that can be generalized across a broader perspective and population. But the quantitative methods are limited in their capacity to capture the meaning of the data, motivations that shape human behavior and organizational experience. It tends to simplify the complexity of what is happening within an organization (Queiros et al., 2017, pp. 381-382).

The qualitative approach, in contrast to quantitative, focuses on deepening the understanding of a given problem by exploring its motivations, and values that cannot be quantified (Queiros et al., 2017, p. 370). Hammersley (2012, p. 1) discusses qualitative research as placing an explanation on words rather than numbers, studying this within a natural setting and looking at how events unfold from the perspective of those involved. The experience of people being close to an event is what gives qualitative research its strength in studies of organizational phenomena where context is decisive (Hammersley, 2021, p. 1213). As with quantitative research, qualitative also has limitations. The findings are more difficult to generalize, the process is time consuming, and the researchers' personal assumptions and perspective on the subject need to be carefully handled (Queiros et al., 2017, pp. 378-379).

This study will adopt a qualitative research method. The research question is not one that can be answered fully with numerical measurement alone. The value created by an AI coding assistant within an organizational context is not fixed, objectively measurable data. It is experienced, interpreted, and perceived differently by different individuals depending on their role in the organization, their past working experience, and what expectations they have towards the implementation of the tool. For the authors to capture this complex situation requires a method that can access the perspective of those involved in their own words.

The primary data collection in this study will be done by semi-structured interviews. Doody & Noonan (2013, p. 2) have identified interviews as the most common method in qualitative research because of the open-ended questions allowing participants to respond in their own words, giving the researchers rich and contextually grounded data. Among the available interview formats, the semi structured approach is the most appropriate for this study.

4.7 Research Design

The main objective when defining the research design is to find the most relevant and reliable plan through which the study can find the most valid findings. A good, determined research design is often characterized by a strong relationship between the chosen methodology, the research question, and the research paradigm for the study (Collis & Hussey, 2021, p. 88). The four most used types of research design are descriptive, exploratory, explanatory, and predictive (Collis & Hussey, 2021, p. 5).

Descriptive research design focuses on the documenting of detailed characteristics of a phenomenon in a precise and systematic way, giving empirical data and evidence for claims (Collis & Hussey, 2021, p. 5). Exploratory research designs look to develop a better understanding of a phenomenon, most typically in areas where previous studies are limited. It looks for patterns and generates ideas rather than testing already studied propositions (Collis & Hussey, 2021, p. 5). Explanatory research adds to the descriptive design by aiming to establish relationships between different variables and understand the connections (Collis & Hussey, 2021, p. 5). The predictive design also adds to the descriptive and explanatory, by using the existing relationships and generalizing the findings and predicting the future outcomes (Collis & Hussey, 2021, p. 5).

This study will adopt a design that combines exploratory and descriptive characteristics, because it directly reflects on the base of the main and sub research questions. The central aim of the research question is to understand where and how the implementation and integration of GitHub Copilot has created value within Vattenfall BA Markets operations department. A question that seeks to map and characterize a phenomenon rather than establishing proof of why it occurred. The study does not set out to test hypotheses or demonstrate a cause-and-effect relationship between GitHub Copilot and specific performance outcomes. Rather, it will try to identify the organizational processes in which financial value can be shown, and to describe how and where that value can be shown in practice.

In this study, the exploratory dimension is primary. As discussed before in the problematizing, peer-reviewed research specifically on this subject, examining the organizational effects of AI tools in real business context is still vague (Ziegler et al., 2022). This study therefore operates in relatively uncharted territory, making exploration an appropriate option for this study to start at. This will also justify the semi-structured interviews as the primary data source, as these allow the authors to follow the participants' experience, rather than constraining the responses to a set of options (Doody & Noonan, 2013, p. 2).

To complement the exploratory, this study will also look at the descriptive approach as it will provide structure to the findings by the exploratory. Once patterns begin to show from the interviews, the study will move towards describing those patterns in a systematic and empirical way. The combination of exploratory and descriptive is well suited to this study's context where the phenomenon is real and ongoing, but the academic frameworks for understanding this are still developing.

4.8 Event Study Within a Case Study

This study will be performed as an event study within a case study, examining Vattenfall BA Markets implementation of GitHub Copilot within the Operations department. A case study is defined by Yin (2018, p. 15) as an empirical inquiry that investigates a contemporary phenomenon in depth and in its real-life context, particularly when the outcomes between this phenomenon and its context are not clear. This design is well suited to this study given that GitHub Copilot’s value creation is hard to separate from the specific organizational, technical, and operational context in which it was deployed. As Flyvbjerg (2006, p. 229) talks about, when the objective is to gather the greatest possible amount of information about a problem or phenomenon, selecting cases based on their information richness is more appropriate than random sampling.

An event study examines whether and how a discrete, identifiable event produces measurable changes in outcomes of interest, by comparing conditions observed before and after the events occurrence (McWilliams & Siegele, 1997, pp. 626-630).

While event studies have traditionally been applied in financial economic to assets markets reactions, the approach has been extended to organizational research to evaluate the effects of strategic interventions such as technology adoptions (Subramani & Walden, 2001, pp. 1-7). In this study, GitHub Copilot's implementation is the event, and the analysis is orientated towards identifying what changes in operational processes, workflows, and perceived value creation outcomes.

The combination of these designs is in alignment with the study's exploratory-descriptive purpose. Collis & Hussey (2021, p. 5) describe exploratory research as suitable where few prior studies exist on a given phenomenon, and where the goal is to identify patterns and develop understanding of the outcomes. The event study structure provides temporal anchoring, while the case study provides the methodological path for collecting qualitative evidence through semi-structured interviews. Together, they enable the study to address both the where and how of value creation from AI implementation.

The case study is appropriate here because of the kind of knowledge it is positioned to create. Yin (2018, p. 15) argues that the design is preferred when the research questions ask how or why about contemporary phenomenon over which the investigation has little or no control; both hold in this study. Under these conditions the case study generates knowledge by preserving the contextual richness that makes the phenomenon intelligible, and the form of generalization it supports is analytic rather than statistical, extending the joint application of the four theories from chapter 3 rather than claiming reproducibility elsewhere (Yin, 2018, pp. 37-38; Flyvbjerg, 2006, p. 229). This sits within the inductive logic adopted in section 4.5 and the pragmatist position in section 4.2-4.3 since the analyses move from particular observations towards broader analytical statements and treats interview accounts and organizational artefacts as evidence relevant to a real-world problem (Saunders et al., 2009, p. 109; Yazan, 2015, pp. 137-139). The method has well documented limitations, primarily concerns about rigor, limited statistical generalizability, and unmanageable volumes (Yin, 2018, pp. 19-21), which the authors mitigate through the structured thematic procedures in section 5.4, by reframing the contribution as analytic and bounding the case in section 8.6, and through purposive sampling and the block structure, researchers proximity is addressed through reflexivity, and the event-study contrast is treated as inferential rather than causal where multiple changes occur simultaneously (Flyvbjerg, 2006, p. 234).

4.9 Usage of Artificial Intelligence

This study has used AI during the research process only for finding sources, feedback on grammar, and discussions of ideas. The AI that has been used are a variety of different generative AIs such as ChatGPT Plus, Microsoft Copilot, and Claude. The authors, together with Vattenfall, have generated ideas for the thesis and have independently written the text. AI has only been used as a support tool and done so aligning with Umeå University's Guidance for the use of AI tools.

AI has not been used to produce drafts or text ready for submission, neither has it been used for producing drafts for rewriting in the author's own words or for combinations of draft into one text. The authors have throughout the whole process been fully responsible for the ideas, analysis, structure, argumentation and wording of the thesis.

Prior to the usage of AI, all the tools have been instructed to help in a way that preserves academic integrity and student independence. Text written by the authors that has been given to an AI tool was done on the instruction to give feedback and ask questions to help improve academic tone and identify weaknesses. The tools have also been used to point out unclear parts in drafts and discussions based on the author's own text and thoughts.

The authors chose to use generative AI tools to improve the grammatical quality of the text, as well as to ensure academic language, by providing the authors feedback on their drafts. The usage of AI was in alignment with the requirement to preserve independence and academic integrity.

4.10 Research References

To make sure of the quality and academic integrity of this study, the authors have used sources that are peer reviewed and accessible through academic databases. It is of very high importance that the literature base in a field such as AI implementation, where a significant volume of publicly available material comes from industry reports, consultancy publications, and practitioner blogs, while they are informative, they do not meet the standard for academic thesis papers. All the literature used to develop the conceptual framework, theoretical framework, and analysis in this study has been taken from peer-reviewed journals or academically recognized publications.

The primary databases in this study were EBSCO Business Source Ultimate, Google Scholar, ScienceDirect, the Umeå University Library catalogue, and DIVA Portal. All these databases provide access to a wide range of peer-reviewed journals within business administration, organizational studies, technology management, and information systems. All of these are relevant to the subject of this thesis. If possible, sources have been accessed in their original published form to avoid misrepresentation.

The literature search was conducted across the databases listed above, using keywords that are based on the subject of the thesis: for the empirical context, GitHub Copilot, AI code assistant, generative AI in software development, AI in finance, AI in energy markets, AI productivity, automated bias, and human-AI collaboration; for the theoretical framework, real options theory, managerial flexibility, bounded rationality, satisficing, task-technology fit, technology adoption, Information Systems Success Model, and productivity paradox; and for the methodology, case study research, event study methodology, thematic analysis, qualitative research design, and pragmatism in business research with Boolean operators and backward and forward citation tracing used to identify additional sources. The literature base was deliberately constructed to extend beyond AI implementation, which on its own would not provide the analytical depth required by the research questions: each of the four theories is supported by foundational sources and more recent application in digital and AI-augmented work, and the methodological literature is drawn from established work in qualitative research design rather than general textbooks. Peer-reviewed scientific articles in established journals and academic papers from recognized publishers are prioritized, while industry and consultancy publications are used selectively and only for empirical context, never to support theoretical or methodological claims, with sources accessed in their original published form wherever possible.

5.0 Practical Method

This chapter covers the practical execution of the research – how data was collected, who was selected to participate in the interviews, how the interviews were structured, what secondary data were drawn from Vattenfall BA Markets, how the materials were analyzed, and how ethical obligations were handled throughout the process. Together these sections give the reader what is needed to evaluate and, if necessary, replicate the empirical work.

5.1 Data Collection

One of the most commonly used data collection tool in qualitative research is the usage of semi-structured interviews because the open-ended format lets the respondents answer in their own words, increasing the possibility of producing rich data that a standard questionnaire cannot generate (Doody & Noonan, 2013, p. 28). As opposed to a fully structured interview, the semi-structured format lets the interviewer follow unexpected threads that might show up while still covering a set of predetermined themes (Doody & Noonan, 2013, p. 28). This flexibility is especially appropriate here due to GitHub Copilot's implementation at Vattenfall BA Markets is recent and still evolving, meaning that some of the most informative responses may come in reply to probing questions that could not have been anticipated at the design stage.

5.2 Selection Rationale

The participants were selected through purposive sampling, a non-probability method in which individuals are chosen because they possess specific knowledge or experience directly relevant to the research question (Collis & Hussey, 2021, p. 192). Random sampling would not serve the purpose here: only employees who actively use GitHub Copilot in financial or operational decision-making contexts have the experiential basis to answer the questions the study poses. The aim was to provide a wide representation across seniority levels and functional roles within Vattenfall BA Markets, rather than depth from a single part of the organization, so that findings reflect the tool's impact across the decision-making chain rather than only at one level.

Initial contact was made by the authors' in order to conduct this study with Vattenfall BA Markets, which provided access to relevant personnel and facilitated scheduling. Participants were then identified in consultation with a contact at the company who had an overview of which departments use GitHub Copilot most intensively. The authors conducted preliminary interviews with approximately 30 individuals at the company to establish an understanding of the usage frequency across departments, which then led to snowball referrals which supplemented this process. No participant was coerced or pressured into participating; all took part on a voluntary basis.

5.2.1 Preliminary Interviews

To determine which employees to conduct semi-structured interviews, the authors conducted approximately 30 preliminary interviews that were about 15 minutes each with employees across Vattenfall BA Markets. The purpose of these preliminary interviews was not to generate empirical material for analysis, but to investigate how frequently and in what context GitHub Copilot was used across Vattenfall BA Markets. This scoping step was necessary to understand the usage intensity, rather than formal title, determines which employees possess the experiential basis required to answer the research questions (Collis & Hussey, 2021, p. 192). Conducting a mapping of these interviews prior to the full interviews is consistent with Saunders et al.'s (2009, p. 237) recommendation that purposive sampling in exploratory case studies benefits from an initial scoping stage to identify information-rich participants.

In these preliminary interviews, snowball referrals emerged organically as respondents pointed the authors towards colleagues with more intensive or distinctive GitHub Copilot usage patterns. Snowball sampling is considered appropriate when the targeted population is difficult to identify through formal organizational structures and when existing participants are well positioned to refer others with relevant experience (Parker, Scott & Geddes, 2019, pp. 1-4). In this study, the technique was supplemented rather than replacing purposive sampling. The initial participants were selected based on department mapping, and referrals were used to extend the sample towards employees whose day-to-day work involves GitHub Copilot to a degree that would otherwise have been invisible from an organizational chart. This combination strengthens the sample information richness, which Flyvbjerg (2006, p. 229) identifies as more important than representativeness in case study research.

All participation was voluntary. No participant was coerced or pressured into taking part, and employees were free to decline the initial conversation or any subsequent interview without consequence. This principle is elaborated further in section 5.5.1 on the informed consent.

5.3 Interview Guide and Process

It is essential to prepare thoroughly and design the interview guides around the research question to conduct an effective semi-structured interview (Doody & Noonan, 2013, p. 28). The guide should function as a navigational tool that supports natural and comfortable interaction with participants while ensuring that all relevant themes are covered systematically (Kallio et al., 2016, pp. 2954-2956). A participant that is relaxed is more likely to provide detailed and reflective accounts of the phenomenon being studied, however, developing such a guide requires anticipating challenges such as how to phrase complex questions to elicit rich and meaningful data (Brinkmann & Kvale, 2015, pp. 149-166).

A recommended approach is to begin interviews with questions that are easy for participants to answer, gradually progressing toward more abstract or cognitively demanding topics (Doody & Noonan, 2013, pp. 29-32; Rubin & Rubin, 2012). In line with this, the interview guide for the study has been structured to open with introductory questions about the participants' role, background, and day-to-day responsibilities. This has allowed participants to settle into the conversation before progressing toward more substantive themes, such as how GitHub Copilot has been integrated into their workflows, perceived changes in decisions being made and any operational or financial implications that have been observed. The guide was adapted slightly depending on the participants' role. Developers were asked questions more oriented towards task-level effects and tool usage, while more senior participants were asked questions leaning more towards organizational outcomes and strategic implications.

The interviews were conducted online and in English. Prior to each interview, participants were informed about the purpose of the study, the voluntary nature of their participation, and their right to withdraw at any point. Subject to participant consent, interviews have been recorded and transcribed to support systematic analysis in line with the ethical considerations outlined in section 5.5.

The interviews were conducted in the manner of four different blocks. The participants are divided into these blocks' dependent on their hierarchical position in the organization.

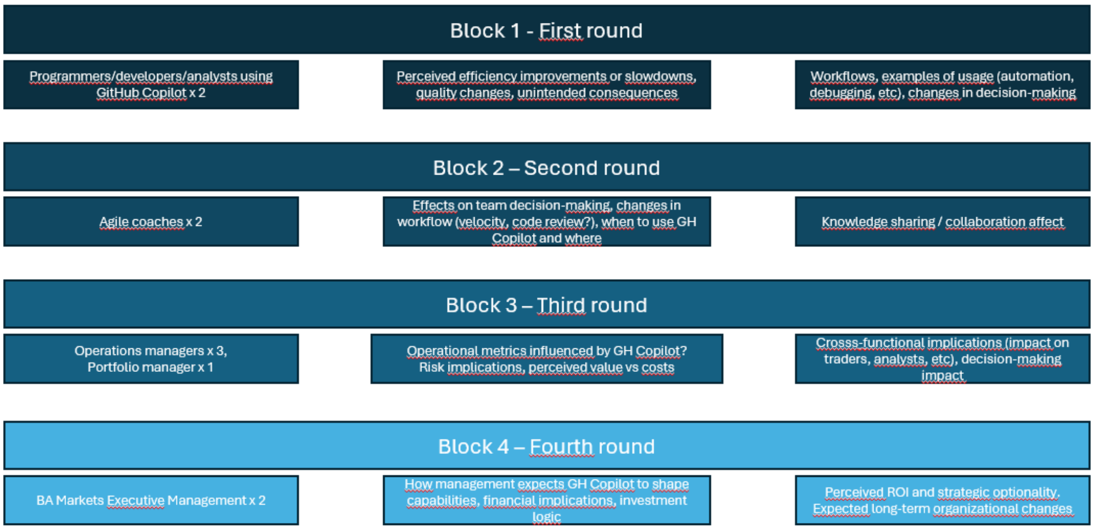

Figure 2. The empirical data collection structure is organized into four sequential interview blocks, each corresponding to a distinct organizational level. Developers, Agile Coaches, Operations and Portfolio Managers, and BA Markets Executive Management. Together with the themes explored at each level. Source: Authors' own construction.

5.4 Data Analysis Method

Yazan (2015) wrote a research paper regarding similarities and differences between the works of Robert Yin, Sharan Merriam, and Robert Stake, in which there are some analytic methods to consider. The analytical methods presented in the research paper by Yazan (2015, pp. 134-152) discuss taking observations apart and reducing the data to examine and categorize in order to make sense of the data. But what is worth highlighting is that Yazan (2015) discusses the importance of finding a method that works for the researcher.

There are several different analysis methods that can be implemented when interviews have been used to gather data which are discussed by Dunwoodie et al. (2023). The different methods are thematic (content), grounded theory, discourse analysis, narrative analysis, and interpretative phenomenological / hermeneutical analysis (Dunwoodie et al., 2023, p. 880). Thematic (content) analysis is an analysis method where the researcher assigns codes to data and structures them into themes. Grounded theory aims to generate new theory directly from data, rather than testing an existing one. Discourse analysis focuses on examining how people use language to construct meaning within their social and organizational contexts, and narrative analysis examines the stories people tell to make sense of their experiences. Lastly, interpretative phenomenological / hermeneutical analysis explores the unique, lived experiences of individuals and how they make sense of those experiences within their contextual environments (Dunwoodie et al., 2023, p. 880). With the information brought up from Yazan, Dunwoodie, and the inductive approach of the thesis, the authors have decided to use the thematic (content) analysis approach.

Thematic analysis was chosen in this study because it offers the flexibility needed to identify patterns across a heterogeneous participant pool while remaining systematic enough to ensure analytical rigor (Braun & Clarke, 2006, pp. 86-93). Unlike grounded theory, which seeks to construct an entirely new theory directly from the data without prior framework (Dunwoodie et al., 2023, p. 880), this study applies four established theories as an analytical lens and uses the empirical material to refine and extend their joint application. Discourse and narrative analysis were considered less suitable, as the research questions concern organizational mechanisms and decision-making behavior rather than language use or personal storytelling (Dunwoodie et al., 2023, p. 880). Interpretive phenomenological analysis was likewise set aside, since the focus lies on collective workflows and operational outcomes rather than individual lived experience (Dunwoodie et al., 2023, p. 880). Thematic analysis therefore provides the most appropriate balance between openness to emergent insight, such as the level problem developed in chapter 7, and structured comparison across participant blocks, allowing the authors to connect the interview material back to the research questions in a consistent manner (Scharp & Sanders, 2019, p. 108).

When using the thematic analysis method, there is a six-step working order created by Braun & Clarke (2006, pp. 86-93) that is applied, which is later brought up by Scharp & Sanders (2019, p. 108). The six-step working order is as follows: (1) becoming familiar with the data, (2) generating coding categories, (3) generating themes, (4) reviewing themes, (5) defining and naming themes, (6) locating exemplars. To become familiar with the data, the researcher needs to look back at transcriptions and/or (re)read the data that was found. It is required to mark interesting features of the data in a systematic approach and then collect the data to generate valid coding categories of higher quality. Generating themes involves bringing together related initial codes and organizing them into broader, meaningful categories. Reviewing themes focuses on evaluating whether these themes accurately reflect the coded extracts and the dataset. Defining and naming the themes means to articulate the core idea each theme represents, clarifying both what it includes and what it excludes. Lastly, locating exemplars involves selecting strong, illustrative quotations or evidence that best demonstrates the theme and clearly connects it to the research question.

The authors began the analysis by thoroughly reviewing the interview recordings and transcripts, as well as any relevant KPIs that participants mentioned or provided as supplementary data. All interviews were conducted in English, and the interview guide was adapted to suit the specific type of interviewee, meaning that certain questions were emphasized or adjusted depending on the interviewee's role and experience. By listening carefully to the recordings and engaging closely with the transcripts and additional materials, the authors became well acquainted with the data set. This process enabled the authors to identify patterns and develop initial coding categories that align with the aim of the thesis. These categories later served as the basis for constructing preliminary themes. The authors then reviewed and refined the themes to ensure that they accurately represented the coded data and reflected the core insights emerging from the interviews and KPI material. Each theme was named according to its central meaning, and finally, the themes were connected back to the research question. This process followed the analytical sequence outlined by Scharp & Sanders and applied a structured thematic approach.

5.5 Ethical Considerations

Ethical considerations are vital to any qualitative research that has human participants, as it creates a role where accurate information can be gathered without causing harm to the individual or organization (Nii Laryeafio & Ogbewe, 2019, p. 95). The ethical section lays out the principles applied throughout the data collection and analysis process in this study.

5.5.1 Informed Consent and Voluntary Participation

All participants in this study were informed about the purpose of the study prior to the interviews. A brief description of the research and interviews was sent out in advance, so participants could make an informed decision if they wanted to participate. Participation was entirely voluntary and in no way was there any pressure or obligation to take part in this study. As Nii Laryeafio & Ogbewe (2019, p. 102) discuss, voluntary participation is essential to be able to ensure that the data collection is honest and reflects genuine responses rather than answers shaped by coercion. Participants were informed prior to, during, and after interviews that they have the right to withdraw at any point of the process without providing a reason, and that data they have provided would be deleted upon a withdrawal.

5.5.2 Confidentiality and Anonymity

For the authors to protect the participants' privacy, all interviews have been anonymized and referred to throughout the thesis as, Participant 1, Participant 2 and so on. The studied organization has been given the option to be anonymized but has given the authors permission to include its name in the study. No identifying personal details such as name, titles beyond general role description, or contact information appear in this thesis. Interview data, including reports and recordings, were only accessible to the authors of this study. This aligns with the principle that collected information must remain confidential and that no third party should have access to the data (Nii Laryeafio & Ogbewe, 2019, p. 102).

5.5.3 GDPR and Anonymity

This study was conducted in compliance with General Data Protection Regulation (GDPR), Regulation (EU) 2016/679. Under article 5 of the GDPR, personal information must be processed lawfully, transparently, and for a specified purpose. The data that has been collected should not be retained longer than what is stated as necessary. In practice, this means that only information relevant to the research question and subject was collected. Participants were not asked to provide personal information as defined under Article 9 GDPR, including data concerning health, ethnicity, or other special categories, as that kind of information is not pertinent to the study's objective. Any accidental inclusion of personal details in interviews was not recorded in a form that could identify the individual. Voice recordings during interviews are stored securely and will be deleted when this thesis is completed. The thesis will be published in alignment with Swedish law (offentlighetsprincipen) and Umeå University's guidelines, and participants were informed of this prior to their participation.

5.5.4 Independence and Research Posture

This study was not commissioned by or written on behalf of Vattenfall. The authors have no employment or financial benefit with Vattenfall that could act as a conflict of interest. During the interviews, actions were taken to avoid leading questions and to allow participants to guide the conversation, in line with the best practice for semi-structured interviewing (Doody & Noonan, 2013). The authors adopt a pragmatist epistemological position, where total objectivity is not the primary concern and the focus is instead on producing research that is practical and useful in answering the research questions (Sauders, Lewis & Thornhill, 2009, p. 109). Nonetheless, efforts were made to ensure that the analysis reflects the participants' own words and interpretations rather than the authors' prior assumptions.

5.5.5 Research Responsibility

Throughout the data collecting and interview phase, the authors have made efforts to avoid situations that could result in discomfort to participants. As Nii Laryeafio & Ogbewe (2019, p. 102) discuss, the researchers have an obligation to eliminate sources that could result in harm or uncomfortable positions for participants. Participants were treated as equals and were given the opportunity to ask questions before, during, and after their interview.

5.6 Ethical work practice

Prior to the data collection, relevant institutional and legal guidelines were reviewed to ensure that the thesis follows applicable regulations and ethical standards. This thesis is in alignment with Umeå University's Thesis manual and Writing on Commission guidelines. But as this thesis is not commissioned, the research was performed independently, therefore ensuring academic integrity and analytical autonomy. The handling and processing of empirical material comply with Swedish law, including the offentlighetsprincipen, as well as GDPR. Ethical considerations related to confidentiality and data protection are explicitly addressed throughout the whole research process.

An initial meeting with the assigned supervisor at Vattenfall BA Markets was conducted to discuss the data handling procedures, confidentiality, and ethical implications related to the interview data collection. After this a description of the study's scope and purpose was sent out to potential participants. Participation in this study was strictly voluntary. Prior to the interviews, participants received an outline of what subjects would be covered and some examples of questions that would be included in the interviews. No personal data is collected during the interviews, ensuring participant anonymity and minimizing privacy risks.

At the start of each interview, participants were again informed about the purpose of the study, their rights as participants, and ethical principles in the research. Consent was obtained for audio recording for transcription purposes. Participation is entirely voluntary, and participants have the right to withdraw at any point without providing justification. If any withdrawal occurs, all collected data are immediately deleted and excluded from the thesis.

The interviews are conducted with efforts to minimize power dynamics and foster a professional and respectful dialogue. The authors aim to avoid situations that may cause discomfort and treat all participants as equal contributors in the thesis. Participants are given the opportunity to ask questions before, during, and after the interviews, as well as to request information regarding the progress of the thesis.

For analytical purposes, interviewees are anonymized and referenced as participant 1, 2, 3…. This approach ensures confidentiality while maintaining analytical rigor and transparency, consistent with ethical standards in finance research.

5.7 Quality criteria

In qualitative case studies, the positivist criteria of validity, reliability, and generalizability translate imperfectly, because of the ontological and epistemological assumption that underlying qualitative inquiry differ from those of quantitative work (Bell et al., 2022, pp. 358-360). Lincoln & Guba (1985, pp. 289-331) brought up an alternative framework based on trustworthiness and organized around credibility, transferability, dependability, confirmability, and authenticity. These criteria align with the pragmatist stance adopted in this study and with the qualitative case-study design and are therefore used here to evaluate the quality of the inquiry.

5.7.1 Credibility

Credibility concerns the degree in which the findings represent the participants' experience accurately (Lincoln & Guba, 1985, p. 296). The study addresses credibility through three mechanisms. First, the four-block design functions as a form of structural triangulation: the implementation of GitHub Copilot at Vattenfall BA Markets is examined from four distinct hierarchical vantage points, allowing convergent and perceptions to be identified and weighed against one another. Second, preliminary scoping interviews refined the interview guide and ensured that the questions targeted experiences relevant to the research question rather than topics the authors assumed to be relevant. Third, all interviews were audio-recorded, transcribed, and analyzed in their original form, reducing the risk that interpretation drifts away from what participants said.

5.7.2 Transferability

Transferability is the qualitative counterpart of generalizability and shifts the burden of judgment to the reader: the researcher's task is to provide a sufficiently rich description of the case context for others to assess whether the findings would apply in another setting (Lincoln & Guba, 1985, p. 316). In order to be able to address this, the study describes Vattenfall BA Markets in detail so that the conditions under which the findings arose are visible to the reader. The study makes no claim to statistical generalizability from a sample of ten interviews in one organization. What it instead offers is an analytical generalizability which takes inspiration of Yin (2018, p. 38): in Chapter 7 the level problem is proposed as a theoretical mechanism that other organizations should examine in their own context, rather than stating it as a universal law that was derived from a single case. The research process that the authors conducted is itself replicable: the multi-level, block-based interview design through which the level problem was identified can be applied by any organization investing in AI to examine whether the same cross-level mediation occurs in its own setting. In this sense, the study offers not only a mechanism to be examined but a procedure for examining it.

5.7.3 Dependability

Dependability is described as whether the research process of the study is documented in a way that would allow another researcher to follow the same steps and reach a defensible finding (Lincoln & Guba, 1985, pp. 316-318). This study follows the six-step thematic analysis produced by Braun & Clarke (2006, pp. 86-93) and has been applied to the material, keeping the interview guides as an appendix, and maintains a correct auditing trail consisting of recordings, transcripts, coding categories, and theme definitions. The interview guides were edited to fit the participants' role but kept consistent for each block, in order to be able to compare answers from similar block participants.

5.7.4 Confirmability

Confirmability addresses the extent to which findings are shaped by the data rather than by researcher bias (Lincoln & Guba, 1985, pp. 318-327). Two threats are particularly relevant to this study. First, the authors have a current working but not commissioned relationship with Vattenfall, which is addressed in section 4.4 and managed in practice by anchoring interpretations explicitly in interview quotations rather than recollection. Second, the four-theory framework introduced in chapter 3 could in principle have predetermined what the authors looked for; this risk mitigated by the inductive coding sequence, in which themes were generated from the empirical material first and theory was then applied to interpret them rather than to filter them. The use of generative AI tools is disclosed in section4.9 and was confined to grammatical feedback on author own written texts.

5.7.5 Authenticity

Authenticity asks whether the researchers fairly represent the multiple perspectives presented in the studied setting (Bell et al., 2022, pp. 361-365). Developers, agile coaches, operational management, and BA Markets executive management perspectives are each represented rather than assembled into a single voice. The findings in chapter 6 are presented block by block, in order for the block-by-block findings to remain visible to the reader, and so that the formatting of the results are not spread out across all blocks simultaneously. Lastly, anonymization protects the studies participants while preserving the substance of their contributions through verbatim quotations taken from the collected transcripts.

5.7.6 Quality Purpose

All these taken together, these criteria are intended to make the conditions under which the study's findings are transparent to the reader, in line with the pragmatist's commitment to usefulness over claims of certainty. The limitations of the design that fall outside the trustworthiness framework are discussed in section 8.6

6.0 Presentation of Findings

This chapter presents the findings gathered through semi-structured interviews with ten employees at Vattenfall BA Markets within the operations department. Rather than reporting individual responses sequentially, the findings have been structured and organized into four analytical blocks that correspond to distinct hierarchical and functional groups within the organization. The structure was designed to allow a layered examination of how GitHub Copilot's effects are perceived and experienced at different levels.

Note: Participant names and precise tenure are anonymized in accordance with the confidentiality agreement described in Section 5.5.

Table 1. Overview of Interview Participants

Participant	Role	Block	Experience	Interview Time and Date
P1	Lead/Fundamental Analyst	Block 1: Programmer & Analyst	Senior	45-60 min 30-03-2026
P2	Developer (Application & Pipelines)	Block 1: Programmer & Analyst	Senior	45-60 min 31-04-2026
P3	Agile Coach	Block 2: Agile Coach	Senior	45-60 min 10-04-2026
P4	Agile Coach	Block 2: Agile Coach	Senior	45-60 min 10-04-2026
P5	Controlling & Strategic Development	Block 3: Management	Senior	45-60 min 14-04-2026
P6	Director of Analysis	Block 3: Management	Senior	45-60 min 14-04-2026
P7	Director Business Infrastructure	Block 3: Management	Senior	45-60 min 14-04-2026
P8	Project Portfolio Manager	Block 3: Management	Senior	45-60 min 15-04-2026
P9	Vice President Controlling & Strategy Development	Block 4: Executive Management	Senior	30-45 min 20-04-2026
P10	Vice President Operations	Block 4: Executive Management	Senior	30-45 min 23-04-2026

6.0.1 Participant Selection and the Rationale for a Block Structure

The selected participants for the study were based on purposive sampling, targeting individuals across four different hierarchical levels within the organization: direct tool users (Block 1), team-process coordinators (Block 2), operational and strategic decision-makers (Block 3), and the BA Markets executive management (Block 4). The reasoning behind this interview structure was to capture how the implementation of GitHub Copilot has changed individual workflows within Block 1, and also how its implementation has affected or failed to affect other organizational outcomes further down the line. Access to the participants was granted through the author's existing relationship with Vattenfall BA Markets, and participants were identified in consultation with a senior internal contact.

Each block consists of two participants, apart from Block 3, which includes four. Within each block, participants have comparable roles and share a common organizational perspective. This allows for internal cross-validation: where participants converge on a finding, that finding carriers greater analytical weight; where they diverge, the divergence itself becomes meaningful. Two participants per block reflects the case study logic of the study, where depth takes precedence over statistical coverage.

Block 3 is the section in the results that stands out most since there are four participants instead of two. This was a conscious decision by the authors as the management layer is internally diverse and all participants occupy distinct positions within the same tier.

Each block in the interview guide has a specific purpose since the participants within each block have separate views and perspectives that are important for the study. Block 1 describes how GitHub Copilot has changed decision-making behavior and task execution; Block 2 looks at whether those changes have a noticeable impact on team workflows and delivery outcomes; Block 3 assesses how management views, characterizes, and controls the tool's operational and financial implications; and Block 4 describes how the investment is perceived and justified at the executive management level.

6.1 Block 1: Programmers and Analysts

The findings in this section are from participants whose primary responsibilities involve hands-on technical work, for example, writing code, building and maintaining quantitative algorithms, developing data pipelines, and producing analytical outputs. The participants within Block 1 represent the group with the most direct and sustained interaction with GitHub Copilot on a day-to-day basis. Their participation offers a detailed perspective on how the tool has changed the nature of technical work within Vattenfall BA Markets, what conditions shape its effectiveness, and where they perceive constraints on its broader impact.

6.1.1 Background and Participant Profile

Prior to the interviews, Participants 1 and 2 were given a short time to introduce themselves and describe the composition of their roles at the organization.

Participant 1 (P1) works as a lead analyst and fundamental analyst, with a primary role within maintaining a market model used to forecast energy prices. In addition to this core function, P1 has also taken on a role supporting colleagues in adopting AI infrastructure projects within the organization, including a data streaming platform and an initiative to improve data discoverability across the business area.

Participant 2 (P2) holds a development-oriented role, focusing on building applications, data pipelines, and services that support operational and analytical functions within BA Markets. P2 described working on complex, multi-component systems and engaging with GitHub
Copilot is a central part of their daily workflow. All participants mentioned that GitHub Copilot has been available to them for a sustained period prior to the interviews, enabling reflection on how the usage had developed over time rather than initial impressions.

6.1.2 Task-Level Decision-Making

A central finding coherent across both participants is that GitHub Copilot has changed how technical work is approached at the task level. The change is not only of speed, although speed improvement has been mentioned, but more on how participants locate information, frame problems, and decide which task to attempt. Both participants described a distinct shift away from the manual, externally dependent workflows that characterized their pre-GitHub Copilot experience.

P1 described the most immediate comparison with Stack Overflow, which is a website that features questions and answers on certain computer programming topics (Stack Overflow, 2026). P1 had previously used the website as the primary source for resolving and consulting technical issues, and questions:

"*Manual work. Before, I was very active on Stack Overflow. Whenever I had a problem, I went there. I also contributed a lot – helping others, things like pricing models, stuff like that. It has changed now: When I have a problem, I simply ask the AI (GitHub Copilot)"*

(P1)

Stack Overflows interactions are intermittent, dependent on community availability and the quality of existing answers and require the user to formulate a query in the absence of their specific codebase context. This is a significant observation beyond the surface, in which one information source has replaced the other. GitHub Copilot, in contrast, operates within the developer's environment and responds to the actual code and context in progress. The structural advantage is one of contextual relevance and immediacy.

P2 has described a similar shift in work processes, noting that GitHub Copilot enabled a complete frontend application which was built in three days instead of the estimated two or three months it would have taken before GitHub Copilot was introduced.

"We just had to build a little front end for one of our applications... usually this would have taken two or three month and would have required one or two developers... With the new tools, I could build this in three days."

(P2)

Both participants have described that the usage of the tool has evolved considerably since the initial launch. P1 described this evolution as three identifiable phases: Code autocompletion at the time of introduction, a chat interface enabling more open-ended interactions, and most recently, agent-based workflows capable of taking over extended task phases.

"We have a lot of coding tasks and pull requests from colleagues to review, and now I even ask Copilot to look at them and give thoughts, then I filter them... In the beginning, everyone looked at code suggestions. Then came the chat interface. Now we have agents that take over tasks."

(P1)

One of the most consistent findings from Block 1 participants is that GitHub Copilot has reduced hesitation around complex or unfamiliar tasks. The participants described situations in which they had previously avoided or delayed certain work due to its technical complexity or their limited prior experience with the required methods, and in which GitHub Copilot had enabled them to proceed effectively.

P1 provided a concrete scenario of this effect in the context of a web scraping task that required authenticated access to a vendor documentation portal:

"Scraping documentation from a vendor's website that required login. I'd never done that before. (GitHub) Copilot created a scraper for me in two hours. Before, that would have taken days".

(P1)

P1 articulated a broader change in how they approach whether to initiate complex work:

"Things I used to hesitate to do because of complexity – I now just do, and results are surprisingly good."

(P1)

P2's account reinforced this pattern from a different technical vantage point. The participant described how GitHub Copilot enabled the exploration and implementations of solutions that would not have previously been attempted:

"It's much more efficient and basically enables or facilitates new business cases. I'm working on a search application that not only uses a lexical search approach, but also semantic search... which is something I wouldn't have done before. In certain aspects, its quite complex."

(P2)

6.1.3 Specification and Verification Work

Alongside changes in how tasks are attempted and approached, participants described a fundamental shift in where the cognitive efforts are directed during technical work. Both participants agreed on the same finding that using GitHub Copilot effectively requires investing significantly more time and attention in defining the problem and the specifications.

"We are still responsible for the code. But I don't see the benefit of learning every detail of every framework. I prefer to invest time into why and what, not how. We only have eight hours."

(P1)

The reflection from P1 discusses the ability to define problems clearly and evaluate the generated outputs from the tool might be more important than the technical implementation of knowledge. P1 continued on how this shapes the actual workflow.

"My workflow is strict: start with requirements and specification. This is where I really look carefully. The planner creates a specification and plan. Often, it's overambitious, so I remove things I don't want – otherwise runtime errors appear later."

(P1)

P2 expressed a related perspective, emphasizing that the quality of GitHub Copilot's output depends on the precision and context provided by the user:

"It depends to a certain extent also how I set it up. Whether I use skills and specify best practices... this increases the reliability quite a lot. Nevertheless, it always has to be challenged, and you always have to look at the results twice."

(P2)

The participants share an interesting view that is significant for the study's research question. The findings themselves represent an ongoing change in decision-making itself, and not only the changes happening in speed or quality. Before the implementation of the tool, a developer would make a continuous series of micro-decisions on code implementation, but as seen here the GitHub Copilot-assisted workflow concentrates decision-making into a distinct upfront phase (the specification), followed by a review and evaluation of the generated output. The restructuring of decision-making in general has implications for both the skill required of technical staff and for the quality of outcomes produced.

6.1.4 Output Quality and Trust

Both participants reported a high level of confidence in the quality of output produced by GitHub Copilot, with an important qualification, quality is contingent on the precision of the specification or context provided in the prompt to GitHub Copilot. The finding emerged consistently and was expressed in similar terms.

P1 compared GitHub Copilot's output quality directly to the work of less experienced colleagues:

"I think it's just better code than like 99% of our colleagues, especially new ones. It follows guidelines quite well, with exceptions. But generally, it's better than what a new colleague would do"

(P1)

GitHub Copilot is not a simple tool requiring extensive supervision, but more so as a virtual colleague that can produce a good enough work that can exceed what less experienced colleagues would generate.

Output quality is closely dependent on the model used and the context provided. P2 described situations where they actively chose different models depending on the complexity and nature of the task at hand, noting that the actual model used is a decisive factor in output quality. In a scenario where the specification is clear and the appropriate model is used, P2 reported reliable, high-quality outputs, in contrast, when the context is incomplete or a wrong model is used, quality and the reliability of the code deteriorate. Within their different tasks, both participants described separate instances where the tool produced hallucinations or lost context during longer interactions, which reinforced the expression of conducting a constant output verification.

6.1.5 Operational and Financial Value

Neither of the participants tried to quantify the financial impact of GitHub Copilot in precise monetary gains, but both provided details that describe concrete operational value creation with clear downstream financial relevance.

P1's most direct relevance to financial value concerned a multi-agent workflow observed in a trading-adjacent team, where a solar generation forecasting model was built within a single working session:

"I saw a great example last week, a team building multi-agent workflows in afternoon. Creating a solar generation model from Germanys registry plus weather forecast. That directly benefit trading. I think this will explode."

(P1)

Within Vattenfall BA Markets energy trading operations, the speed at which forecasting models or quantitative models can be developed and deployed can give direct commercial significance. The accuracy of these models are a large factor for the firms trading performance in renewable energy markets, where positions are taken based on expected generation output. In this case, their ability to create a working forecast model within hours rather than weeks translates into competitive advantage.

P2's account for operational and financial value creation was based on the enablement of new capabilities, rather than the acceleration of existing ones. The frontend application was built in three days versus an estimated two to three months, and the search application that would not have been attempted at all, was completed with the usage of GitHub Copilot. This represents not only marginal productivity gains but order-of-magnitude reductions in labor input.

P1 also described a non-linear productivity trajectory, small initial gains at the time of adoption followed by substantially larger effects as both the tools capabilities and participants prompting skills developed simultaneously.

"Much higher productivity. In the beginning, small improvements. Now huge. I write long prompts with full detail, often GitHub Copilot solves it in one shot."

(P1)

In the authors discussion with P2 they mentioned an estimated productivity increase of between thirty and forty percent for developer tasks, while acknowledging that the figure is an approximation and depends on the proportion of development work in any given period. This finding is relevant regarding how the financial value of the investment in GitHub Copilot could be assessed. Using a one-time evaluation conducted shortly after the initial deployment would only capture the early phases of the curve and would substantially undershoot the value generated as user proficiency increases over time.

6.1.6 Governance and Structural Constraints

Other than the operational value described in the previous section, the participants identified structural conditions that constrain how GitHub Copilot's potential is realized within the organization. Most notably, P1 and P2 offered strikingly divergent assessments of the organizational governance framework surrounding the tool's use.

P1 described the formal organizational support of GitHub Copilot critically:

"Organizational support is zero. It was a long struggle to get more premium requests. People hit the 300-request limit quickly. Without premium, you fall back to simpler models, which are not good. No additional guidelines either."

(P1)

P1 identifies two specific structural constraints on value realization. First, access limitations: standard users are restricted to a lower capability model, meaning that the productivity gains described in section 6.1.2 through 6.1.5 are only available to those who have obtained premium access. Secondly, the absence of training or formal guidelines means that the quality of GitHub Copilot usage is highly dependent on individual initiative and self-directed learning, producing uneven outcomes across the use of population.

P1 also gave the description of uncertainty around data governance, whether Vattenfall had contractual agreements in place to prevent GitHub from using proprietary code for model training purposes.

"I hope we have agreements where GitHub doesn't use out code for model training."

(**P1)**

The use of "hope" rather than certainty indicates that the terms governing GitHub Copilot's use had not been clearly communicated to the participant. In an energy trading context, where code may contain commercially sensitive or market-related logic, this represents a meaningful operational risk.

P2's perception of the organizational governance framework was in contrast,

"I think it couldn't be better actually. Its well supported and I always feel that regarding these types of technologies, we are kind of the front runners. Particularly also when it comes to big companies."

(P2)

The difference between P1 and P2 in this dimension is one of the most significant findings in Block 1. Both participants work within the same organization and broadly within the same business area, yet their experiences differ substantially. They may reflect differences in team context, proactive governance-seeking behavior, or expectations regarding adequate governance. Both participants converged, however, on training as the primary lever through which the organization could improve outcomes from its GitHub Copilot investment.

"Invest more in training. Access to GitHub Copilot is good, but many don't know how to use it efficiently. We need upskilling."

(P1)

P2 similarly emphasized data findability and accessibility as the foundational requirement for maximizing AI tool value, recommending that Vattenfall invest in improving data quality, classification, and discoverability as the basis upon which downstream AI-powered services can be built.

6.1.7 Summary of Block 1 Findings

Block 1 showcases that the implementation and usage of GitHub Copilot have changed how technical employees at Vattenfall BA Markets approach their work. The tool has shifted the information seeking away from external community sources which was common practice within their role and moved towards a context-aware assistant embedded directly within their working environment. The implementation has also reduced overall hesitation surrounding complex tasks as well as expanding the scope of work employees/practitioners are willing to undertake. In addition to this, it has restructured decision-making by having users concentrate into upfront specifications rather than the manual implementation of the code. GitHub Copilot has also effectively produced output quality that participants in this block saw as “exceeding” the baseline of less experienced colleagues, contingent on the precision of the input provided. These task-level changes have generated concrete operational value, including order-of-magnitude reductions in development time and the enablement of entirely new capabilities. Having said that, the value is currently being constrained by uneven governance, the absence of structured training, as well as the support that the users receive for using the tool in practice.

6.2 Block 2: Agile Coaches

The findings in this section are taken from two participants whose roles center on assisting team effectiveness, managing transformation processes, and coaching agile teams at Vattenfall BA Markets’ Operations function. Unlike the participants in Block 1, the agile coaches do not interact with GitHub Copilot as a development tool in their own workflows. Their perspective gives insight into how the tool’s effects manifest at the team level instead of the individual level, such as, workflow dynamics, collaboration patterns, and delivery processes. This difference makes their observations relevant for understanding whether the observed task-level productivity gains seen in Block 1 translate into broader operational outcomes.

6.2.1 Background and Participant Profiles

Participant 3 (P3) and Participant 4 (P4) are both agile coaches within Vattenfall BA Markets. P3 has been at Vattenfall for several years and described the role as being able to cover several functions including the training on agile-team frameworks, cross-team coordination, and being a mediator during times of conflict. P4 has a background in organizational transformation and during the interview described the coaching role as helping teams work together to achieve a higher performance. In total, there are currently five agile coaches within the organization supporting approximately thirty-two teams across software development, algorithmic trading, and platform operations.

Neither participant comes from a technical background, and both stated that GitHub Copilot had not been previously mentioned during their conversations with the different teams. When P4 asked their team(s) directly whether they were using the tool, the response was immediate:

"I asked them: is (are) any of you using GitHub Copilot? The answer was like – all of us."

(P4)

The response from the team members alone indicates that GitHub Copilot had been implemented universally among the developers, yet, largely without visibility at the team coordination level.

6.2.2 Productivity Gains and Delivery Speed

A remarkable finding from the agile coaches is that while output has increased among the Block 1 participants, the end-to-end delivery process has not become faster. Both participants reflected upon a bottleneck forming at the quality assurance and review stages of the software development pipeline. P4 described observing an increase in the volume of pull requests (which is how you formally propose changes to a software project instead of messing with the original code directly), which can be an indicator that code is being produced at a higher rate than before (Caballar & Stryker, 2026).

"What I have noticed is that in the teams there are a lot more pull request, and requests for others to review. So, at some point the people have become a bit of a bottleneck because I think that the code is being created faster, but then we're still asking people to review. It's a lot more in the pipeline than maybe half a year ago."

(P4)

P3 adds to this observation across the teams in which they are involved, describing testing and review as the single largest constraint in the delivery pipeline:

"There's much more pull request bottlenecks on testing and reviews. So this is my current observation, the biggest bottleneck in the process, end to end, is QA (Quality Assurance) and testing. It is the one that takes the longest."

(P3)

The development phase seems to have increased, but the overall delivery speed has not improved at the same pace due to constraints in downstream processes. P3 mentioned the development and usage of automated testing but noted that it has not been a viable solution for this issue. Vattenfall operates critical energy infrastructure, where security requirements and the volume of edge cases in systems that must run continuously make automated end-to-end testing difficult to implement:

"We are working with sensitive things. They need to be super stable. IT security is a huge topic… we cannot automate a lot of testing as there are so many edge cases, so many exceptions at the moment."

(P3)

6.2.3 Adoption Friction and Shared Norms

An operational consequence of GitHub Copilot's adoption that was described by the Block 2 participants were the increases in team-level friction arising from the absence of shared norms around how AI-generated code is produced, reviewed, and integrated. This friction is prominent when individual developers use GitHub Copilot intensively without a team-level agreement on standards. P3 described a specific case involving newly onboarded developers who used GitHub Copilot to produce a large quantity of code rapidly. While the speed of output appeared productive, it created more work for the rest of the team.

"He uses AI to write a lot of code and then builds a mock-up, which is great because prototyping becomes much quicker. But then the team needs to follow up – because if you don't feed it well, if you don't have an agent that knows all the guidelines and the coding standards, it creates a lot of code that needs to be understood and maybe refactored. So, it can create a lot of extra effort, initially."

(P3)

P4 discussed this dynamic in relation to interpersonal tension within the team:

"It creates another kind of team dynamic, and it starts to create tension as well. I experience that it's like one person is screaming – they have too many things to be reviewed, but we have agreed that we can wait five days."

(P4)

Regarding recently onboarded team members, the participants note that they appear to reach a productive level faster than they previously did and attributed this partly to the availability of GitHub Copilot and AI tools. P4 described new hires delivering valuable code within a month of joining the organization – something that was not common before. However, the same case can show the limitation of speed without proper alignment: faster code produced by a new hire created review burdens that offset part of the onboarding gain for the team.

6.2.4 Implementation Decisions and Operational Outcomes

The decision to adopt GitHub Copilot was made centrally at the Vattenfall Group level, and no formal training program was introduced specifically for GitHub Copilot. P3 noted that mandatory training had been provided for M365 Copilot, but not for GitHub Copilot. Developers can install and use it without formal onboarding:

"From what I know, people can install it and use it, but there is no structured training around. That's a potential opportunity, I believe."

(P3)

In the absence of formal guidance, P4 described how knowledge about effective GitHub Copilot usage has developed organically, through peer support within teams and through broader AI community channels. Both participants categorized this as "bottom-up" and informal rather than systematically managed.

"AI is also introducing a change that people need to cope with, which means that change needs to be managed. You can introduce Copilot top-down and then expect people to figure it out themselves – but eventually you need to manage it and try to fence it in a bit:

what can we keep from what we've been experimented with? And what should we stop doing? Because otherwise this will lead to friction between people"

(P3)

6.2.5 Summary of Block 2 Findings

The presented Block 2 findings give insight into the operational implications of GitHub Copilot, and that it extends beyond the individual task level documented in Block 1, and that those implications have not been transitioned cleanly into improvements at the team and process level. The accelerated code production that Block 1 participants described are also observed by the Block 2 participants, but these gains are currently absorbed by a bottleneck that has formed at the quality assurance and code review stage. Although pull requests have increased (which can be a metric to track code production), the review capacity has not, and due to Vattenfall's operational environment places a cap on what can be relieved through an automated testing agent. The result of this is that the output an individual can produce has increased whilst the delivery of working software has not, implying that there is an imbalance in the current production pipeline.

Adoption occurred largely below the coordination layer of the organization, neither coach was aware that GitHub Copilot use was effectively universal in their teams until they raised the question directly; and the absence of shared norms around AI-generated work has produced team-level friction which can be seen by the rapidly generated mock-up code, mounting review queues, and disagreements over standards. A partial counterweight sits alongside this friction, in that newly onboarded developers reach productive contribution faster than before – though whether that upstream gain is realized depends on the same capacity that the bottleneck constrains.

6.3 Block 3: Operations Head and Functional Managers

The findings in this section are drawn from four participants whose roles place them in resource-allocation and governance positions within Vattenfall BA Markets. Exposure to GitHub Copilot within this block is mixed in two dimensions: some participants have used the tool themselves, and some have not, and some lead teams in which the tool is widely adopted while others have no direct reports who use it. Their perspectives address both components of the research question directly: the decision-making behavior surrounding the implementation at the managerial level, and the conditions under which its financial implications are, or are not, currently assessed within the organization.

6.3.1 Background and Participant Profiles

Participant 5 (P5) has a senior financial and operational oversight role within BA Markets, in which their responsibilities span across forecasting costs, recruitment, and the management of operations cost structure. P5 has direct visibility into how GitHub Copilot is recognized within the IT cost-charge model and into the standard project-portfolio investment methodology.

Participant 6 (P6) is the head of an analysis department within BA Markets and manages teams which include both developers and quantitative analysts, like the participants who were interviewed in Block 1. P6 has also experimented with AI coding tools and has taken bottom-up actions to encourage adoption within the department, this includes internal communications and support for in-person sessions facilitated by the organization's AI coaches.

Participant 7 (P7) is a functional manager within Operations with line responsibility across several development and asset-efficiency teams, including teams operating continuous shift coverage. P7's vantage spans multiple team archetypes and is characterized less by hands-on use than by exposure to the consequences of adoption for code review, testing, and developer-learning dynamics.

Participant 8 (P8) is BA Markets' project portfolio manager, responsible for the processes through which significant development initiatives are evaluated. P8 coordinates a cross-BA forum that aligns proposed initiatives with strategy and expected benefits, making the perspective portfolio-level and end-to-end.

6.3.2 Basis for the Investment Decision

The current investment in GitHub Copilot is justified on the basis of expected value and professional judgement rather than empirical measurement. None of the four participants in this Block could point to a specific formal evaluation, return-on-investment calculation, or structured performance assessment that has been conducted or isolated specifically for GitHub Copilot.

P7 articulated this directly, framing the investment rationale in terms of belief:

"It's based on belief. We do not have scientifically good data for the performance of a software developer nor for the performance of a team developing software. So it's not like we have a baseline and then we say, hey, they will speed up by X percent."

(P7)

P7 explained that the cost of GitHub Copilot is sufficiently low relative to developer salaries that the investment threshold for formal justification has not been reached. At approximately 40 EUR per month per user for the standard tier, the tool costs the equivalent of one to two hours of developer time – a level at which even modest productivity improvements render the investment self-evidently worthwhile:

"Our investment is, if you would measure them by their opportunity cost, we're talking about an hour a month or two hours a month that we invest into tooling."

(P7)

P5 described a similar outlook from the financial oversight perspective. With total IT budget costs in the range of 35 million EUR and the Copilot expenditure embedded within a much larger Microsoft contract, the tool's cost is effectively invisible in the broader cost structure:

"I think we see this more as enablers in the business. I think we are quite far from scrutinising the efficiency of it, which perhaps is also not ideal – because there is another side to efficiency: that we could get even more out of it if we invested more, for example in training."

(P5)

P6 would also confirm this framing, stating that the current organizational posture is one of enablement instead of cost control. The priority at this stage is to encourage adoption rather than restrict it.

“Right now, the cost is at such a low level… For now, let’s worry about that people actually use this stuff. But then at some point, we need to restrict them.”

(P6)

P8, from a portfolio management perspective, noted that the investment cannot be directly traced to specific project outcomes. The effect of GitHub Copilot, if present, is embedded in the general “noise” of multiple concurrent improvements in processes, structures, and team maturity:

“I have not seen anything traceable or relatable to GitHub. I believe the productivity is continuously improving. But from my angle, there are a lot of factors coming in.”

(P8)

6.3.3 Managerial Awareness of the Bottleneck

When presented with the findings from Block 2 regarding the QA and code review bottleneck, the Block 3 participants responded with varying degrees of recognition. The bottleneck brought up by the agile coaches had not been formally documented at the management level, but several participants acknowledged its plausibility and offered structural explanations.

P7 had heard similar reports from a subset of teams, particularly in the “asset efficiency” area, and distinguished between two separate constraints: code review delays caused by understaffed teams, and QA delays caused by the scarcity of end users available for testing. In particular, shift-based dispatch operations where personnel operate on a 24/7 rotation with minimal slack:

“Almost none of our teams have testers. So, it’s not that we have dedicated testing teams in our development flows. When they needing someone to QA my feature, in most cases they are talking about end users that are available for end user tests.”

(P7)

P5 had not observed the bottleneck but mentioned a structural interpretation of the findings: if development velocity has increased while the review and testing functions remain staffed at levels calibrated to the pre-Copilot pace, the bottleneck is a predictable consequence of misaligned resource allocation.

“It sounds like the review function is staffed based on the previous pace of development… So, long-term, you could need to rethink the competency distribution or the resource allocation.”

(P5)

P6 noted that measuring output increases from a managerial position is rather difficult because IT work complexity is hard to judge from the outside looking in:

“Can I tell from my position? It’s very hard, because the problem is in IT work and analysis work, how long something takes is very difficult to judge from the outside.”

(P6)

P8, on the other hand, confirmed from the portfolio perspective that testing delays are recognizable constraints on time-to-market, though the participant could not attribute this specifically to increased development velocity from GitHub Copilot:

"What I would say, it is probably time to market, because this is something I do observe that there is bottlenecks on resources when it comes to testing."

(P8)

6.3.4 Governance Approach

The Block 3 participants brought up the deliberate decision to keep governance around GitHub Copilot minimal at the current stage of adoption. The rationale being that the organization is in an acceleration phase where the priority is to build adoption and generate experience rather than to control or standardize usage.

P7 articulated this position clearly:

"In general, I would say the least thing that we need are rules around this at the moment. We are in the acceleration phase of using it. We are trying to inspire. We are trying to make it as easily usable as possible."

(P7)

This is a conscious strategy, described by P6 as comparable to the earlier experience with cloud cost management: an initial period of unrestricted encouragement followed by a later phase of cost management once usage patterns are established:

"The policy right now is, look, we're at a point where we need people to be doing these things. We shouldn't scare them off by restricted them financially."

(P6)

6.3.5 Value Visibility at the Managerial Level

A theme reoccurring throughout all Block 3 interviews is that the value that is generated by GitHub Copilot, most prominent at the task level, does not propagate throughout the firm in a form that is visible or usable for managerial decision-making. The current state (invisibility of the effects caused by the implementation of GitHub Copilot) affect portfolio prioritization, the justification for the investment, and capacity planning.

P8 described this dynamic in detail, noting that even if productivity improvements exist, they are embedded within a complex system of concurrent changes that makes attribution to any single tool essentially impossible:

"Whether the decision would have changed, I could not say. The decision quality would definitely improve. There is no doubt about the security that you have, because what we do is we take decisions under uncertainty."

(P8)

Absence of a visible value does not mean the absence of value itself, but it reduces the confidence and defensibility of investment decisions. In a portfolio setting, where numerous projects compete for limited funds, a technology whose advantages are difficult to pinpoint is at a disadvantage in comparison to projects with clear business justifications.

6.3.6 Effects on Competency and Team Composition

The participants in Block 3 discussed early signs that GitHub Copilot as well as other AI tools are starting to impact the team compositions and the profiles that recruiters look at. The impacts are results of the tool's overall adoption at the organizational level, even though it is still evolving and not yet systematically managed.

In a scenario where P7 was determining the size of a specialist UI pool at the end of the previous year, P7 chose to staff the team at a lower headcount than the previous standard, with the explicit expectation that AI tools such as Figma AI and GitHub Copilot would compensate for the reduced capacity:

"This is where I took the decision to low-ball that. But equip them, for example, with Figma AI and the right tools."

(P7)

When discussing recruitment patterns, P5 mentioned that they had observed a broader shift in recruitment patterns. The organization is leaning towards hiring junior profiles and candidates with non-traditional backgrounds, this being partly because AI tools reduce the need for a deeper expertise in specific programming languages or frameworks:

"I notice it is happening quite a lot with competency profiles... We are also bringing in many more juniors, for example, and one reason is that they don't have fixed tools or a preferred programming language."

(P5)

P7 provided a nuanced counterpoint regarding junior developers. While acknowledging that newcomers may adopt AI tools faster, P7 also described cases where juniors produced working code with AI assistance but could not explain the logic of what they had built.

"We also have juniors where we had doubt that they could even understand what they have built at the end of the day. They are doing brilliant in AI. But if they cannot explain me even on a higher functional level whet they have just built, they're also failing the job."

(P7)

P6 expressed a similar forward-looking worry: the topic of what human roles remain becomes important as AI takes over more code-writing and possibly code-reviewing tasks. According to P6, the response varies according on how vital the system is; algorithmic trading necessitates strong human monitoring, yet less critical applications might be able to tolerate a less human role.

"If I'm doing all automated algo trading then I need really strong defences. If I'm building something where a human can later on catch the errors by using it, its less critical."

(P6)

6.3.7 Data Readiness as a Constraint

During the interviews, two participants identified data quality and data accessibility as the current constraint on the future value that AI tools can deliver within Vattenfall BA Markets. This concern extends beyond GitHub Copilot specifically and addresses the infrastructure upon which all AI-powered capabilities depend on.

During the interview with P7, the authors were shown a Model Context Protocol (MCP) integration in which the user can write natural language queries against the organization's market data storage. While the prototype worked for simple queries, P7 mentioned that more complex or ambiguous questions asked would simply fail. This was due to the overall quality of the metadata stored in their dataset.

"I'm not afraid that we will have too few people picking up software development. I'm afraid that the AI will be limited, the value will be limited by not having data accessible in a good way."

(P7)

P2 in Block 1 had independently identified the same constraint, recommending that Vattenfall invest in data findability, classification, and quality as the necessary foundation for downstream AI-powered services.

6.3.8 Summary of Block 3 Findings

The operational value that had been documented in Block 1 and 2 is mostly invisible at the managerial and strategic levels of the organization. Currently, there appears to be no formal evaluation system in place, and the decision to invest in GitHub Copilot is justified based on belief and a low relative cost rather than performance statistics. The quality assurance and code review bottleneck, which was identified in Block 2, is acknowledge as plausible by Block 3 participants but had not surfaced or been addressed prior to the authors interviews with them. Governance over the usage of the tool is deliberately light and is seen as a strategic choice to prioritize adoption over control in its current stage. However, several participants recognized that this approach may be leaving significant value unrealized. The early effects on team composition and competence requirements are currently emerging, with managers beginning to adjust headcounts and profiles in anticipation of AI augmented productivity. Although, the absence of measurement mechanisms means that these adjustments are based on judgement rather than on data. Finally, data readiness was identified as the primary constraint on future value creation, a concern that comes together with findings from Block 1 and points towards a structural investment requirement that extends beyond the tool itself.

6.4 Block 4: BA Markets Executive Management

Block 4 represents findings from interviews with two participants occupying the most senior positions in the organizational hierarchy examined in this study. Participant 9 (P9) holds a senior management role with responsibility for financial oversight of the organization's development portfolio, while participant 10 (P10) is head of operations for Vattenfall BA Markets, with cross-functional responsibility for software delivery capacity across multiple business units. Both occupy positions at several organizational levels above the development teams and agile coaches' interviews in the preceding blocks, and neither is involved in day-to-day engineering work. Their perspectives offer the broadest vantage point in this study: how – if at all – the effect of GitHub Copilot have propagated upwards to strategic and financial decision-making.

6.4.1 Strategic Invisibility at the Leadership Level

A clear theme that arose during the interviews with the Block 4 participants was the absence of any organic awareness of GitHub Copilot at the BA Markets executive management level. When P9 was asked directly whether changes attributable to the tool had reached their level, their answer was unambiguous:

"No, not at all. I wouldn't have known about it if you hadn't told me."

(P9)

The absence was described as a reflection of the tool's existing visibility within the company rather than a lack of communication. When the authors asked P9 whether there would be any noticeable aftereffects from an abrupt discontinuation of the investment in GitHub Copilot from their point of view, the participant backed the same point:

"No. I didn't even know about it before you started digging into this."

(P9)

While more directly involved in promoting adoption across business units, P10 similarly acknowledged that the effect had not been formally elevated to strategic discussions. The usage varies significantly across various business units; in some cases, the adoption is near-universal whilst in other areas only a small percentage use it. Although this is the case, the overall impact remains unmeasured at the organizational level:

"*We are not clearly measuring this, but I can see that teams that are rather advanced are also having bigger effects in terms of productivity."*

(P10)

The overall observation that the value from the investment is perceived but not quantified is a pattern identified across earlier block interviews. A frequent subject matter in previous block interviews is the general opinion that the investment's value is felt but not measured. Because of its low license cost relative to the overall budget, participants in Block 3 described the tool as financially invisible. However, Block 4 reveals an additional layer of invisibility: the operational effects themselves have not translated into any signals that reach the organization's strategic level.

6.4.2 The Testing Bottleneck

The emergence of a quality assurance and code review bottleneck was first mentioned by the agile coaches in Block 2, even if that is the case, both participants in this block independently discussed the bottleneck. For P10, the forming of a bottleneck had already shown some visibility from their point of view. When asked whether asked about the bottleneck, and if it were observable from the participant's position, P10 confirmed:

"How this shows up on my radar is that I know projects where we are not making progress because testing is delayed. I would say I see more cases where we are running into issues in testing capacity"

(P10)

At the same time, P10 framed the bottleneck as potentially transitional rather than structural. The participant made a distinction between the current mismatch in which development velocity has increased but testing methods remain unchanged, and a future state in which AI-assisted testing absorbs some of the workload:

"If you start becoming fast in development because AI, and then you're still doing testing the classical way, you have of course a bit of mismatch in your capabilities. But that's still to be seen whether this is a temporary thing."

(P10)

On the other hand, P9 engaged with the bottleneck from a portfolio management perspective. When the concept was brought up during the interview, the participant recognized its strategic implications and proposed a structural approach: rather than expanding testing capacity alongside development, the organization might reconfigure existing roles – allowing GitHub Copilot to support code generation while developers are redeployed to quality assurance functions:

"Maybe we don't need the coders to do the actual coding. Instead, we could put more coders on testing. They can read code and verify it…"

(P9)

P9 acknowledged this as a temporary reasoning, noting that the key precondition of a formal bottleneck analysis within the portfolio function had not yet been carried out:

"It is clear that analyzing bottlenecks is something we need to do. We don't have control over that right now."

(P9)

As mentioned previously, the bottleneck identified in earlier blocks has become partially visible at the most senior level examined in this study. But, neither participant in Block 4 had a systematic response prepared for solving the issue and deferred to future analysis as a prerequisite for any action being taken.

6.4.3 Investment as Enablement

The current framing of the investment rationale for GitHub Copilot is informal due to the low licensing cost which had exempted the tool from any structured financial evaluation. Participants in Block 4 knew of this observation and gave a more detailed description of the investment philosophy driving AI-enabled capabilities.

When asked, P10 explained the current view on the investment:

"We are not looking at this as a cost-saving target. We are really looking at this as an enablement. I'm looking at AI as an efficiency tool. We haven't done this step of "hey, bring in AI and now reduce workforce by 20%". I would rather see 20% more outcome."

(P10)

GitHub Copilot is currently positioned to increase throughput within the existing resource base, not to justify a future reduction in the organization's workforce. P10 elaborated on the long-term strategic rationale by linking GitHub Copilot to a broader view of AI's role in a data-intensive organization:

"We are a data driven company. We have weather data, allocation data, market data – lots and lots of data. We want to make decisions that don't need to be 100% right, but just more often right than wrong. And that's pretty much a description of what you can expect from AI."

(P10)

P10 discussed AI's overall involvement in trading and market decisions as a long-term strategic vision, and GitHub Copilots worth in software development being the more immediate return. From a budgeting perspective, P9 noted that the organization's development portfolio has historically operated under resource constraints which often lead to prioritization trade-offs:

"We have too little money to develop everything we want. That is why we have a prioritization, a development portfolio. We always have to deprioritize some very good projects that would deliver significant value to the whole of Vattenfall."

(P9)

The possibility of cutting development time per initiative, even modestly, has the potential to carry direct portfolio implications. P9 expressed interest in understanding how many portfolio initiatives could benefit from GitHub Copilot-assisted development, and by how much the time-to-completion could be reduced.

6.4.4 Organizational Design: Licenses Versus People

Beyond the investment rationale, participants within this block identified a structural question: as AI-enabled development matures and expands, what is the appropriate balance between licensing costs and human capital? The question has been largely absent from earlier blocks and emerges at the BA Markets executive management level as an organizational design problem that may occur in the future.

P9 mentioned this tension:

"Organizational setup is very important here, and it comes down to licences versus employees. What is the cost of not giving people these tools? What does a person who can use the tools cost? They may be more expensive, but what is the output?"

(P9)

The cost comparison represents a different framing from the typical cost-saving logic which is associated with AI adoption. P9 anticipated that this issue would become particularly prominent during hiring decisions, and especially at points of natural staff turnover:

"This analysis will become very important for us and will affect how we make decisions, especially around new recruitment when people leave – what is sometimes called attrition."

(P9)

The participant raised the possibility of a structural shift in the organization's cost base: as license costs replace some proportion of personnel costs, the balance between operating expenditure and capital expenditure may change.

6.4.5 Measurement as a Prerequisite for Action

The current absence of a proficient measurement instrument for the implementation of GitHub Copilot emerged as the main constraint on BA Markets executive management's ability to act on the findings surfaced in this study. Both participants acknowledged that the implementation of the tool had affected development speed, portfolio throughput, and cost structures were real in principle but invisible in practice. P10 was candid about the difficulty of the measurement problem, distinguishing between the conceptual clarity of the value proposition and the empirical difficulty of demonstrating it:

"Measuring productivity of software development is never an easy task. What's the complexity? How fast should you be? There are methodologies about function points, but they're not really helpful."

(P10)

The participant indicated that the organization was at an early stage of developing measurement approaches, and had raised the question internally, the previous day about when to introduce usage targets:

"I think we need a couple of projects where we can make this effect more tangible. For example, a front-end heavy development, doing a questionnaire afterwards: what was our initial estimate, how fast did we actually deliver? And see what kind of percentages we come up with."

(P10)

P10 also noted that consistent usage was a prerequisite for meaningful team-level measurement: given that individual developers were currently using the tool in quite different ways, aggregating productivity effects across a team was not yet feasible.

P9, approaching the same constraint from the portfolio level, identified bottleneck analysis as the entry point for measurement:

"This is something I might now require from the portfolio function going forward – that we continuously track which bottlenecks we have, because we have never really had that conversation before."

(P9)

In the interviews with both participants, it became apparent that without any reliable measurement for GitHub Copilot, no prospective planning tool could emerge. Without any measurable effect, or evaluation of the investment in the tool, a conversation about adjusting delivery expectations, rebalancing resource allocation, or scaling the investment could not meaningfully begin.

6.4.6 Financial Implications Beyond Software Development

Another dimension that extends the financial implications of Gen-AI considerably further than the preceding sub-sections on GitHub Copilot's effects within software development workflows was discussed with P10. Beyond the acceleration of code delivery, the participant discussed the potential that AI carries in the organization's core commercial function: energy trading and market decision-making. The argument was based on the nature of trading operations as it is a high-volume, data-intensive environment where the quality of decisions is measured probabilistically rather than by a certain outcome:

"We want to make decisions that don't need to be 100% right, but just more often right than wrong. And that's pretty much a description of what you can expect from AI."

(P10)

The probabilistic output of LLMs, which is often cited as a limitation in contexts that require precision, is instead reframed here as a natural fit for trading environments, as having an edge in the market lies not in uncertainty but in marginal accuracy.

"In trading, you just need to be more often right than you are wrong. Which is, if you think about what AI provides, it's not a certainty that things are right, but if you ask the question the right way, if you provide the right data, you have a higher likelihood that things are right. And that is the reason why I believe in decision making, it will help us a lot."

(P10)

P10 distinguished this from the GitHub Copilot use case specifically, framing it instead as a longer-term strategic horizon for AI within BA Markets:

"If you look at use cases of efficiency in software development and becoming smarter in the decisions you make, those are really two overarching themes for AI in energy markets, and I think it would be a pity if we wouldn't explore these"

(P10)

AI is not yet being used to analyze and process weather, market, and operational data to make trading and allocation decisions. If it were, the impact on organizational performance could be significant: improvements in decision accuracy within energy markets can lead to substantial financial value of a magnitude that would dwarf the software development productivity gains mentioned throughout this study. But, as there is no current framework in place to isolate or quantify the effect of GitHub Copilot, the financial significance of further AI adoption at Vattenfall cannot be adequately assessed by sheer code velocity alone. The larger financial gains may reside in the domains where AI augments the judgement of commercial decision-makers, not just the throughput of software delivery teams.

6.4.7 Summary of Block 4 Findings

The Block 4 findings reveal that the awareness of GitHub Copilot within the executive management of BA Markets was somewhat split. One participant was somewhat aware of its operational effects prior to the interviews, while the other was not and expected no real effect would register if the implementation were discontinued. The bottleneck that was earlier identified was confirmed to be partially visible at this level, with P10 noting project delays due to testing capabilities and P9 proposing a structural redeployment of developer roles in response. The rationale for the investment was framed explicitly as an enablement to gain more output from their current resources rather than simply reducing headcount. The participants also identified a forward-looking organizational design, regarding the balance between license costs and human capital. A further extension or possibility of financial implications beyond software development was brought up by P10. But to pursue that venture, the current investment in GitHub Copilot needs to be measured correctly to produce enough evidence for an additional investment. Altogether, the measurement gap remains the largest constraint on any strategic action across all themes identified in this block.

7.0 Analysis and Discussion

This chapter will interpret the empirical findings presented in chapter 6 through the analytical lens introduced in chapter 3. Instead of applying each theory in isolation across the data, the chapter will organize the analysis around four themes that emerge from the empirical material itself, with theory applied within each theme to explain the observed patterns. The chapter will give a deeper understanding of the patterns and the meaning behind the data presented in chapter 6.

7.0.1 Table of Summarized Findings

The table below structures the output of the thematic analysis described in section 5.4, applied to the interview material presented in Chapter 6. Four themes emerge from analyzing the ten interviews across the four participant blocks, each capturing a distinct dimension of how GitHub Copilot's implementation is affecting Vattenfall BA Markets. For each theme, the table states the main finding and indicates which component of the research question it primarily addresses; each theme is then discussed in its own section below.

Table 2 Thematic analysis

Theme finding	RQ Component	Main Finding
1. Adoption & Individual fit	Sub RQ, Assessability	Substantial individual productivity gains: Approx 30%-40% on developer's tasks, with order of magnitude. reductions on specific applications
	Main RQ, Behavioral	Developers now spend more time defining what to build and less time writing the code itself.
	Main RQ, Behavioral	Developers learn the tool on their own, as there was no formal training, and views in organizational support diverge.
2. Workflow disruption & bottlenecks	Main RQ, Behavioral	Testing and code review have become the bottleneck in the delivery.
	Main RQ, Behavioral	Faster coding has not led to faster delivery.
	Main RQ, Behavioral	Agile coaches did not know everyone on their teams was using GitHub Copilot.
3. Decision-making under uncertainty	Sub RQ, Assessability	The investment is justified by belief, not measurement; no baseline or evaluation framework exists.
	Sub RQ, Assessability	Governance is deliberately light to encourage adoption, but there are no instruments to track results.
	Main RQ, Behavioral	Managers acknowledge the bottleneck but have not formally addressed it.
4. Financial & Operational implications	Sub RQ, Assessability	The license cost is too small to scrutinize; GitHub Copilot is treated as enablement, not a cost-saving tool.
	Sub RQ, Assessability	Productivity gain cannot be traced to financial or delivery outcomes but hiring and team composition are already shifting.
	Sub RQ, Assessability	Data quality and accessibility, not the tool itself, will limit further value
Cross block analysis		Effects do not add up across all block levels; individual gain, team workflows, and management decisions are disconnected.

7.1 Analysis of Findings

This section explains the four themes summarized in the table above; each theme presented in its own subsection. The analysis stays close to the empirical material, drawing on interview quotations and patterns observed across the four participant blocks. The aim here is to establish what the data shows about each theme before turning *to the theoretical frameworks. The themes are presented in the order in which they emerged from the thematic coding: technology adoption and individual fit, followed by workflow disruption and emergent bottlenecks, decision-making uncertainty, and finally the financial and operational implications.*

7.1.1 Technology Adoption and Individual Fit

The most reoccurring theme discussed during the interviews with participants from Block 1 was how the implementation of GitHub Copilot has altered time spent on a specific task and the way they approach their work. Both participants gave examples of different tasks where an productivity gain is obvious. P1 described building a frontend application in just three days, that would have previous taken two-three months. Similar to what P2 described by building a working web scraper in just 2 hours that would have taken several days before. At this point these changes in time spent on a task are not marginal productivity gain anymore. They have become order-or-magnitude reductions in time to output.

The change is not only about efficiency, as Block 1 participants describe a deeper shift in how technical work is done. Where developers prior to GitHub Copilot spent most of their time writing and debugging code, they now spend more time defining the problem upfront and reviewing what the tool produces. P1 explained this directly: “Time is invested in the why and the what, not the how”. P2 makes the same point from a different angle; the reliability of GitHub Copilot’s output depends on how well the prompt is set up. The cognitive work has not disappeared; it has moved from execution or code writing towards explanation for a new code and reviewing the new code.

A second pattern is that the shift is based on self-learning, and not something that has been taught, as neither participant received nor went through any structured training. They developed their use of patterns through trial and error and informal exchange with colleagues. The agile coaches in Block 2 confirm this from outside; GitHub Copilot adoption was universal among their developers, but no formal training program exists, and knowledge spread through bottom-up rather than top-down. P1 and P2 also diverge sharply on whether the organization supports the tool well. P1 described support as essentially absent; P2 described Vattenfall as a front-runner. Yet they both work in the same organization. This difference is informative: in the absence of structured support, the experience of feeling supported depends on what each developer happens to discover or be told.

A third observation that links the pattern above is that the amount of productivity gain depends on the developer´s skill to explain and understand the problem at hand, not just the tool. Where the task is clearly defined and the right model is selected, the output is reliable and the gains can be huge. Where the explanation to the tool is unclear or the wrong model is used, the output is unreliable and the gains evaporate. The benefit of the tool, in other words, is unevenly distributed even among developers who have access to it, because some have developed the higher skill level that the tool depends on, and others have not. P1’s productivity trajectory illustrates this compounding effect directly. Small initial gains at the beginning of the adoption, followed by substantially larger effects as both the tools' capabilities and the user’s prompting skills developed in parallel. A one-time evaluation conducted shortly after deployment would capture only the early phase of this curve.

A fourth observation the authors made, which is less obvious but just as important, is the concern for trust in the tool. Block 1 participants describe high confidence in the tool's output, P1 notes that the code is better than what most new colleagues produce, but they also describe explicit verification practices. P2 states that results always must be challenged and looked at twice. Both participants described situations where the tool produced hallucinations or lost context in longer interactions. Developers have learnt by trial and error when the tool is reliable and when it's not. Therefore, trust in GitHub Copilot has become conditional dependent on the code produced passing the users own built verification routines rather than just blindly accepting it. This also gives the study an additional clear piece of evidence that what upstream skill differentiates the better skilled users from the rest. This is also subject that the organization has not provided formal training in.

The key finding is that GitHub Copilot has produced real, sometimes large, individual productivity gains. The gains come with a shift in cognitive work towards instructions and explanations, and that they are unevenly distributed because adoption of the tool was self-directed, and the organization provided no structured training.

7.1.2 Workflow Disruption and Emergent Bottlenecks

Block 2 introduced evidence that Block 1 participants alone could not have seen. The agile coaches do not use GitHub Copilot themselves and have no stake in defending its value. But their observation that code and pull requests volume has increased is consistent and in line with what Block 1 participants have described about productivity gains. This has resulted in the formation of a bottleneck in the quality assurance and code review stage in the process. During the interviews P3 described this observed bottleneck as the largest constraint and time absorbed in the end-to-end delivery process. Discussions about automated quality assurance and code review were brought up during the interviews. But because of the current operational context including critical infrastructures, IT security requirements, and a high volume of edge cases, Block 2 participants are operating in; automated testing becomes restricted and cannot relive the increase pressure.

The first layer of the pipeline in end-to-end delivery process has accelerated exponentially, but the layers after it have not had the same journey. This has resulted in a bottleneck formation with the same output as before rather than a higher product delivery rate. The process as a whole has become rather uneven than accelerated. This is the most important operational finding in the study, because it directly contradicts the assumption that individual productivity gains will translate into organizational performance gains.

During the interviews two additional patterns emerged and became obvious to the authors. The pattern that stood out mostly to the authors was that the participants did not seem to have an understanding of how GitHub Copilot worked, and that it was universally adopted within their teams. Until they had asked team members directly. Upon asking this, P4 described asking this question and getting a response for the whole team that everyone is using the tool. This was to the authors and to this study significant in itself, team leaders responsible for team coordination had little to no knowledge about a tool universally used in the teams with a direct effect on workflow and code review burden. Adoption had travelled below their visibility line. Second, P3 and P4 describe team-level friction emerging from the absence of shared norms, one developer producing large quantities of mock-up code that the rest of the team then had to refactor, tension over review queues building up, and disagreements over standards. The friction is not caused by the tool itself. It is caused by adoption running ahead of the team's ability to coordinate around it.

A more constructive finding sits alongside the friction. Both coaches report that newly onboarded developers reach a productive level faster than was previously the case, and they attribute part of this to the availability of GitHub Copilot. P4 describes new hires delivering valuable code within a month of joining the organization, a pattern that was rare prior to the implementation. The same evidence that documents downstream friction therefore also documents an upstream gain, the on-ramp into productive contributions has shortened. The two effects offset each other partially, with the net direction depending on whether the team has the review capacity to absorb the faster arriving output. Where it does, the gain is real; where it does not, the onboarding speed simply moves the bottleneck closer to the start of the queue.

The obvious finding the authors could see here is that increased productivity from Block 1 does not mean a faster end-to-end delivery time. Because Block 2 is not as technical advanced in the AI coding assistance space and in change of the quality assurance and code review stage, this becomes the main time consumer and constraint for the end-to-end delivery. The adoption of GitHub Copilot in the developer space has developed faster than the downstream possibility to absorb its increased output and also the organization's ability to visibility see it.

7.1.3 Decision-Making Uncertainty: Governance and Management Response

Blocks 3 and 4 describe how managers and BA Markets executive management made the original decision to adopt GitHub Copilot and how they govern its use now. Three patterns in particular stand out. The first is that the investment is justified by belief rather than by measurement. P7 states this directly: there is no scientifically valid baseline against which to measure the effect of the tool in teams, and the decision to invest is based on professional judgement. P5 could also confirm this by explaining the cost structure of the tool. Licenses costs are approximately forty euros per month per user and since the cost is considered to be very small for a large organization like Vattenfall, it falls below the cost ceiling that would trigger an evaluation of the tool.

The second pattern is that governance is deliberately light. P7 describes the organization as in an acceleration phase where formal rules would be counterproductive. P6 connects this to a prior pattern with cloud cost management: an initial period of unrestricted encouragement followed by a later phase of cost discipline once usage matures. This is a strategic decision and should not be seen as negligence. The meaning and message that P10 states reinforce the idea that GitHub Copilot should be treated as an enablement. The aim of this second pattern is to increase production by about twenty percent more output, instead of decreasing their workforce by twenty percent. The third pattern is the perception gap. The bottleneck that Block 2 identified clearly had not been formally documented at the management level prior to the interviews.

When presented with the finding, P5 read it as a predictable consequence of misaligned resource allocation, P7 recognized it in specific areas, and P10 confirmed it was visible at the leadership level in delayed projects. P9 understood the issue and gave their idea why this is. That the identification of the bottleneck could not have been done, nor an analysis of the problem that would justify a reallocation of resources. The information about the bottleneck could not and have not travelled within the organization in a way that would prompt managerial action needed. The three patterns identified are connected. Light governance keeps the cost of being wrong low, which is sensible under uncertainty. But light governance combined with no measurement infrastructure means the organization cannot tell whether the early-stage posture is still appropriate or whether the time has come to act on what the data shows. The bottleneck is the operational symptom of this gap.

A separate dimension of managerial decision-making appears in P10's framing of AI's role in the broader trading context. The participant described Vattenfall BA Markets as a data-driven company whose competitive advantage lies in making decisions that are more often right than wrong, not in being one hundred percent right. This framing reads the probabilistic output of AI tools, often treated as a limitation in contexts requiring precision, as a natural fit for trading decisions where marginal accuracy is what produces commercial performance. This is a distinct argument from the productivity case for GitHub Copilot in software development, and it points toward a longer-term strategic horizon in which AI tools support trading judgement directly. The argument is not currently supported by evidence in the empirical material – no interview produced concrete examples of AI support for trading decisions – but it is part of how BA Markets executive management frames the investment, and it bears on how decisions about further AI investment are likely to be justified.

The overall adoption and governance of GitHub Copilot is coherent as a low-commitment strategy under uncertainty, but the organization has not built the measurement infrastructure that would allow it to act on what the data is currently showing.

The overview of adoption and governance is well reasoned and consistent with a low commitment strategy based on belief. This strategy is also in line with the fact that the organization has not built the measurement tools that would enable them to make decisions based on the data it might show. But this is also their biggest constraint in the evolvement of this journey. As a result of this, the bottleneck can only be acknowledged but not addressed since no one can measure it or act on it.

7.1.4 Financial and Operational Implications

This sub-section will address the sub-research question more directly. Looking at what extent, the financial implication of GitHub Copilot can be assessed within Vattenfall Ba Markets, and does this answer reveal. The empirical material produces a structured answer in three parts.

Firstly, the cost side of this investment. This is essentially invisible. The license for GitHub Copilot is very small individually and absorbed within a larger Microsoft enterprise contract, falling below the cost ceiling that would trigger a financial evaluation. Secondly, the productivity gains from GitHub Copilot have remained at the individual developer level (Block 1) and are currently not visible in the end-to-end delivery or financial outcomes. Because of the bottleneck formed at the QA and code-review stage (Block 2) that absorbs these gains prior to reaching the overall and total measurement. Third, decisions that anticipate financial value are already being made based on beliefs rather than measured evidence.

Two further pieces of evidence matter. The first is that hiring and team composition are already shifting in anticipation of AI-augmented productivity. P7 explained how they staffed a UI specialist team with a reduced headcount than the previous standard. Because of the expectation that AI tools they are using will make up for this.

P5 reports a broader move towards junior hires and non-traditional technical backgrounds. These are operational consequences of the tool that have begun to happen in resource decisions, even though no measurement basis exists for them. The decisions are based on the same belief that justifies the investment itself. P7 also raises a counterpoint that complicates this trajectory: some junior developers using AI tools produce working code but cannot explain at a higher functional level what they have built. The competency shift is therefore not a clean substitution of cheaper labor for more expensive labor, it depends on whether the human role being preserved is one of code production or one of system understanding, and the answer to that is not yet settled organizationally.

The second is the data readiness constraint. P7 and P2 who have different positions in the organization and were interviewed separately identify the same ceiling on future value. Quality and accessibility of organizational data, including metadata and discoverability, will determine how far AI-augmented workflows can go. This is a constraint that further license investment alone cannot lift.

P9 makes one further argument that bears directly on the financial case. The development portfolio at Vattenfall BA Markets operates under resource constraints that forces prioritization and not every initiative judged valuable can be funded. Even a modest compression of development time per initiative, if it could be reliably attributed to Copilot, would translate into more initiatives moving through the portfolio in the same period. P9 expressed interest in understanding how many portfolio initiatives could benefit from GitHub Copilot-assisted development, and by how much time-to-completion could be reduced but acknowledged that the analysis does not exist. This is the financial case for measurement, articulated by a portfolio decision-maker, not because the cost of the tool needs justifying, but because portfolio-level decisions cannot be improved without it.

All these pieces describe a value pattern that is real but uneven. The value is positive at the individual level, where the gains are documented. It is ambiguous at the organizational level, because the bottleneck currently absorbs the gains and no measurement infrastructure exists to detect what gets through. From a strategic positioning standpoint, the situation is mixed. On the positive side, the organization has acquired practitioner familiarity and has begun to develop quantifiable competencies. On the risk side, it is making early commitments – in areas like hiring and team structure – based on beliefs rather than on evidence that the value is being realized. The main finding here is that GitHub Copilot is value-positive at the individual level, ambiguous at the organizational level on current instruments, and value-positive in strategic terms subject to the qualification that the strategic value is held rather than measured. Hiring and team composition are shifting based on belief, and data readiness is the next structural constraint on what further investment can deliver.

7.2 Theoretical Reflection

This section revisits the four themes with the theoretical framework from Chapter 3. Each subsection takes one theory, identifies what it explains in the empirical material, and *notes where it falls short or has to be extended. The section is deliberately compact: the empirical work has been done above, and theory is used here as an analytical tool, not as a separate body of content.*

7.2.1 Real Options Theory

Myers (1977, pp. 147-155) and Dixit & Pindyck (1994, pp. 17-34) frame an investment under uncertainty as the purchase of a right rather than an obligation: the firm pays a small premium today in exchange for the option to expand, extend, or abandon as conditions evolve. Read against this framework, the Vattenfall BA Markets approach to GitHub Copilot displays clear option logic. The license cost is the option premium, and the light governance preserves flexibility. P10's explicit framing of the investment as enablement rather than cost reduction is option language made explicit. Li et al. (2025) extend the framework to technology-intensive investments where outcomes are highly uncertain, and the case fits that profile well.

The framework also surfaces a tension. Options have value only if they can be exercised at the right moment, and optimal exercise requires monitoring the conditions under which exercise becomes most attractive or appropriate. Theme 3 documented the absence of that monitoring. The organization has acquired the option but lacks the instruments needed to know when to act. The application here is conceptual rather than quantitative, option values are not calculated, and this is the limitation flagged in section 3.5 about the qualitative use of ROT in this study.

There are three specific options that are shown by the empirical material. The first one is an option for capability scaling. By developing a user's experience and knowledge of AI-assisted development, the organization can in practice delegate and direct a user towards more advanced tooling. This includes agent-based workflows that P1 describes, without having to bear the entire cost of a cold-start adoption

The second option is related to competency and reallocation, where the human capital portfolio is being directed into AI-augmented designs. The hiring and team composition changes reported in theme 4 are early examples of this option. The third option is to restructure the current workforce around the new production frontier. P9s rationale for transferring developer to code review and quality assurance was expressed during the interview rather than as part of a form procedure. All three options have been acquired; only the first two have been partially exercised, and none are being managed with the analytical apparatus that ROT assumes.

7.2.2 Bounded Rationality

According to Simon (1955), actors adopt the first solution that meets a workable limit rather than searching extensively for the best one, describing decision-making under information and cognitive constraints as satisficing. Working similarly to the 80- 20 method, also known as the Pareto principle. In the first theme, developers are satisficed with use patterns through trial-and-error in the absence of training. In the second theme, teams are satisficed with the integration of AI-generated work through emergent practice rather than explicit governance and guidelines. Lastly, managers were satisficed with the investment evaluation because the cost was too low to justify formal scrutiny. Each of these themes is locally rational; the information needed to do better was not available at the level where the decision was being made.

The framework also helps to explain the perception gap that runs through Themes 1 to 3. Developers, coaches, and managers each have access to the information visible from their personal vantage point. Cross-level information gathering requires deliberate investment in feedback loops that the organization has not yet developed. The level problem (developed in Section 7.6), is, at its root, a bounded rationality problem: each layer of the organization is satisficing its own information, and no mechanism is currently lifting that information upwards throughout the organization. The limitation mentioned in Section 3.5 is also visible – the usage of GitHub Copilot may introduce new cognitive dependencies as well as relieving old ones – and the verification practices both Block 1 participants describe are developers to manage that risk.

7.2.3 Task-Technology Fit

Goodhue & Thompson (1995) treat fit as the alignment between a technology's capabilities and the requirements of the tasks it is applied to. The empirical material supports a high-fit verdict for a substantial portion of the developers' work, and a low-fit verdict for tasks that exceed the model's representational capacity. This could be when long interaction with a lot of context happens, the model loses understanding of the complex domain-specific task at hand and starts to suggest hallucinated suggestions.

Two extensions of the framework are needed. First, the "fit" is co-produced; the original framework treats "fit" as a property of the technology-task pair, but Theme 1 shows that "fit" also depends on the developer's specification skill. Two developers working on the same nominal task will produce different levels of "fit" if one invests in upstream definition and the other does not. The second extension is that "fit" operates at multiple levels. Theme 2 shows that a configuration in which "fit" at the individual level is high while "fit" at the process level is low: the tool changed the production rate without changing the capacity of the downstream stages, and the pipeline became imbalanced. Neither extension is present in the original formulation, and both are empirically necessary for generative tools.

7.2.4 Information Systems Success Model

According to DeLone & McLean (2003), there is a chain where information and system quality influence user happiness and use, which in turn generate net benefits for both individuals and organizations. The chain runs strongly through individual impact. System quality and information quality are reported as high, but conditional on input precision. Use is universal, satisfaction is high, and individual impact is high. The chain therefore predicts large organizational benefits.

The empirical material shows that this prediction is partially blocked. Theme 2 shows that the translation from individual impact to organizational impact is mediated by downstream processes whose own fit has not been adjusted. The model can accommodate this by reading net benefits as the residual after process-level mediation, but it does not predict that the mediation will be substantial without complementary investment. This connects directly to Brynjolfsson, Rock & Syverson's (2017) productivity-paradox argument, in which the gap between AI capability and aggregate productivity is mediated by complementary investments organizations have not yet made. In Vattenfall BA Markets case, it displays the major complementary investments (structured training, governance norms, measurement infrastructure, data readiness) as either non-existent or underdeveloped, and Corrado, Haskel and Jona-Lasinio (2021) provide further explanation that the value accumulates in intangibles that conventional accounting does not capture, which matches with Theme 4 directly. Toma's (2025) explanation of the gap between AI investment and evidence of return applies in full.

7.3 How the Theories Connect, and the Level Problem

The four theories form a chain that runs through the empirical material in a definite order. Real Option Theory explains why the organization invested at all under uncertainty, the cost was small, the option was potentially large, and the flexibility was preserved. Bounded Rationality explains how the investment was decided and how it has been governed, through satisficing at each organizational layer, with information trapped at the level where it was generated. Task-Technology Fit explains where the investment produces value and where it does not, at the individual task level the fit is high and conditional on user effort, at the process level the fit has broken down. Information Systems Success Model explains why the value visible at the individual level does not currently aggregate to the organizational level, the chain that the model describes is mediated by complementary investments that have not been made.

The most consistent observation across the four themes is that effects do not aggregate cleanly across organizational levels. Individual gains do not become team-level delivery acceleration. Team-level adoption does not become managerial information. Managerial adoption does not produce option-exercising infrastructures. Each transition between levels is mediated by processes, instruments, or feedback loops that the organization has not yet built. This is the level problem, and it is a finding in its own right.

Each theory in the framework is most informative at a particular level. ROT operates at the strategic investment level. Bounded Rationality operates at the level of individual decisionmakers under constraints. TTF operates at the user-task (with the extension developed above for multilevel fit). The IS Success Model operates at the level of organizational outcomes. Using the four together provides the ability to trace effects across the level transitions that any single theory would treat as relatively transparent. No single theory in the framework would have surfaced the level problem, because each one stops where another one begins. The combined application is what makes the problem visible.

The level problem also explains why the financial argument in this study is necessarily inferential rather than directly demonstrated. Each level of the organization can produce evidence about its own behavior; developers can describe their productivity gains, coaches can describe their delivery bottleneck, managers can describe their investment logic, and BA Markets executive management can describe its strategic posture. What none of them can produce, on the current measurement infrastructure, is evidence about how these layers interact together. The financial implications of the implementation live precisely in those interactions, in whether individuals' gains survive the bottleneck, in whether managerial decisions are calibrated to actuals rather than imagined outcomes, in whether the option position is being exercised at value-maximizing moments. This is the answer to the sub-question stated earlier in the paper. Financial implications cannot currently be assessed at Vattenfall BA Markets because the effects of the implementation are mediated at the transitions between organizational levels, and the organization has not yet built the instruments that would make those mediations visible. The level problem is therefore not merely an explanation of why financial assessment is presently impossible – it is a structural claim about what would have to be true for such an assessment to be possible at all. The honest answer to the financial dimension of the research question is that the evidence required to settle it does not yet exist within the organization.

7.4 Chapter summary

This chapter has interpreted the empirical material from Chapter 6 through the four-theory framework introduced in Chapter 3. There are four interconnected themes that surfaced in the analysis: a bottleneck at the quality assurance and code review; governance and investment decisions made through satisficing and option-preservation logic; and the invisibility of the current financial implications shown from the investment. The connection of the four themes is the level problem: The value that is generated at one organizational level does not transfer to the next, as the organization has not built in the instruments or infrastructure that would make those transitions visible. Chapter 8 develops these results into answers for the research questions and into practical recommendations.

8. Conclusion

This chapter presents the principal findings of the study, separates the theoretical contributions from practical recommendations, addresses ethical and societal implications, reflects on the limitation of the design, and indicates direction for future research. It compresses the argument developed in Chapter 7 into a clear answer to the two-tier research question – first the main question on decision-making behavior and operational implications, then the sub-question on the assessability of financial implications.

8.1 Answering the Research Question

The purpose of this study was to provide a deeper understanding of how the implementation of GitHub Copilot has influenced organizational behavior and operational outcomes within a real organizational context, and to identify the mechanisms that explain the observed effects. The study posed one main research question on decision-making behavior and operational implications, and one sub-question on the extent to which the financial implications of the implementation can currently be assessed. The chapter answers each in turn and then reflects on what the sub-question's answer reveals.

8.1.1 Main RQ: Behavioral and Operational Findings

The main research question asked how the implementation of GitHub Copilot has influenced decision-making behavior at Vattenfall BA Markets, and what operational implications this carries for the organization. The study answers this question in two separate dimensions.

Within the behavioral dimension of the research question, the study provides insights on how the implementation of GitHub Copilot has influenced decision-making at four organizational levels. Within the developer level (Block 1), the work seems to have shifted from programming to specifying tasks upfront and verifying the generated output. At the team level (Block 2), emergent practices have driven the decisions on how to integrate AI-generated work rather than any explicit governance or norms, and the layer responsible for delivery flow only became aware of a universal GitHub Copilot adoption after asking team members directly. At the managerial level (Block 3), the investment and governance posture have been made under low-information conditions that might not be typical for investment decisions. Although having said that, the license cost was small enough to be under the threshold for any traditional evaluation, and the investment has been maintained on professional judgement rather than on measured outcome data. At the strategic level (Block 4), BA Markets executive management has framed the investment in options terms, as an enablement for the employees in order for the organization to expand its knowledge within the area and possibly benefit from it. However, further investments would need strong proof-of-concept, as the monitoring infrastructure is lacking, which halters the "option" to be exercised at a value-maximizing moment.

Addressing the second part of the main research question, which lies in the operational dimension, the study shows three key findings. First, the individual productivity gains are real and substantial. Block 1, participants reported large reductions in time spent on specific task types, as well as an approximate thirty to forty percent improvement on developer tasks more generally. Second, there is an apparent bottleneck in the quality assurance and code-review stage which has become the binding constraint. Individual gains that the developers/analysts gain do not translate into a faster end-to-end delivery, which creates a situation where the organization misses out on the value capture of the investment. Third, there are already clear signs that the implementation is already producing operational consequences in both upstream and downstream process of code production. Agile coaches were unaware of the usage of GitHub Copilot, hiring patterns are shifting slowly in anticipation of AI-augmented productivity, and team compositions could be staffed below the previous standard headcount on the expectation that AI tools would compensate.

8.1.2 Sub-RQ: Assessability of Financial Implications

The sub-question asked to what extent the financial implications of the implementation can be assessed within the organization, and what the answer to that question reveals about how generative AI investments should be evaluated. The empirical material produces a structured answer in three parts.

First, the cost side of the investment is structurally invisible. The GitHub Copilot license is small in absolute terms and absorbed within a broader Microsoft enterprise contract, falling below the threshold at which formal financial evaluation is triggered. Second, the benefit side is quarantined at the individual level: real and measurable productivity gains at the developer level are not currently traceable through delivery or financial outcomes, because the QA and code-review bottleneck absorbs them before they reach aggregate measurement. Third, decisions that anticipate financial value are already being made on the basis of belief rather than measured evidence – most visibly in hiring and team composition, but also in the strategic posture that frames the investment in option terms while lacking the instruments that would allow the option position to be exercised at value-maximizing moments.

The sub-question's answer is therefore that the financial implications of the implementation are not currently assessable at Vattenfall BA Markets. This is not a confession of empirical limitation; it is a structured finding about the conditions under which such an assessment would be possible.

8.1.3 What the Sub-RQ Answer Reveals

The organization is unable to assess its financial implications under current organizational conditions is itself a revealing answer. The study points to three structural features of the case that generalize beyond Vattenfall BA Markets. The first is the level problem: it explains how the effects do not aggregate cleanly across the organizational levels, and each transition between levels is mediated by processes, or feedback loops that the organization has not yet built. The financial implications of the implementation lie precisely in those mediations. The second is the absence of measurement infrastructure as a binding rather than incidental constraint – the organization cannot exercise its option position at value-maximizing moments because it has no instruments to know when those moments occur. Third is that AI-tool ROI debates more broadly cannot be resolved at any single organizational level, because the value being debated lives in the cross-level interactions that single-level evaluation cannot see. The sub-question's answer is therefore not only a finding about Vattenfall BA Markets; it is an empirical demonstration of why the financial evaluation question for generative AI tools needs to be asked differently than it currently is.

8.2 Key Conclusions

Five conclusions, drawn from the four themes of the analysis and the cross-cutting level problem, summarize what the study has shown. Each finding is tagged with the research question component it primarily addresses.

The first one, GitHub Copilot has produced real and in some cases large individual productivity gains, and these gains come with a shift in the skill set of developer work toward problem framing and quality assurance (Main RQ - Behavioral). The gains have become unevenly distributed across developers, not because they have access to different tools, but because of differences in input quality, model selection, and quality assurance routines that the organization has not formally taught their developers or Agile coaches during or after the introduction to GitHub Copilot. The difference in productivity gains has therefore become a structural issue of the implementation, and not just some accidental variations.

Second, those individual gains do not currently translate into faster end-to-end delivery of working software (Main RQ – Operational). Code volume and pull request volume have grown; the quality assurance and code review stage have not, and a bottleneck has formed there. Automated testing cannot fully relieve the pressure given the operational context of critical infrastructure and IT security requirements. Adoption travelled below the visibility line of the team-coordination layer of the organization, with the agile coaches responsible for delivery flow only learning of universal GitHub Copilot use after asking directly. Adoption has therefore outpaced both the downstream capacity to absorb its output and the organization's ability to see it.

Third, both the original investment in GitHub Copilot and its subsequent governance reflect distinct but reinforcing logic at the managerial and strategic levels (Main RQ – Behavioral). At the managerial level, decisions have been made by satisficing under low-information conditions; at the strategic level, the same decisions have been framed in option-preservation terms, preserving flexibility under genuine uncertainty about how the technology will evolve and where its value will materialize. The light touch is internally coherent across both levels, but it lacks the monitoring infrastructure that would allow the option to be exercised at value-maximizing moments. The organization holds an option position without the instruments required to manage it.

Fourth, the financial implications of the implementation are not currently assessable within the organization (Sub-RQ). Cost is essentially invisible embedded in a broader Microsoft enterprise contract. The benefits from the individual level are stuck because of the QA bottleneck. Because decisions that anticipate financial value are being made on the basis of belief rather than measured evidence. Data quality and accessibility, not the tool itself, will bound any further value capture. The unassessability is structural, not incidental, and is the substantive answer to the sub-question.

Fifth, value from the implementation is unevenly distributed across the four organizational levels: visible at the individual level, visible but not actionable at the agile-team level, ambiguous at the management level, and latent in option form at the strategic level (Both RQs). Each layer of the organization can give information about its own behavior, but no organizational layer can give information into a view of how the layers interact with each other. The authors identify this as the "level problem". It addresses both research questions: it explains why behavioral and operational change does not travel cleanly across levels (Main RQ), and it identifies the structural reason why financial implications are unassessable (Sub-RQ).

8.3 Theoretical Contribution

The study contributes with three concrete theoretical contributions, each tied to a specific gap in the existing literature, instead of just making a general claim about why the topic matters.

The first concerns the level problem itself. The study does not propose the level problem as a generic claim that organizations are layered, which would be uncontroversial. It proposes the level problem as a methodological claim about AI-tool evaluation: the effects of generative AI tools cannot be assessed at any single organizational level because the effects are mediated at the transitions between levels, and the mediations are themselves the substantive phenomenon. No single theory in the framework used by this study surfaces the level problem, but combining the four theories does. This is a contribution to how AI-investment/implementation research should be framed methodologically, not only to what it should report. It also responds directly to the call made by Brynjolfsson, Rock, and Syverson (2019) and Brynjolfsson et al. (2019) research that goes beyond simple productivity measures and investigates the organizational mechanisms by which AI creates tangible value. In doing so, it provides a direct theoretical answer to the sub-question of this study: the assessability of financial implications is itself a function of organizational architecture, and the conditions under which AI-tool ROI can be evaluated are themselves a research object that prior frameworks have not isolated.

The second contribution is the progression of the Task-Technology Fit framework. Goodhue and Thompson (1995) treat fit as a property of the technology-task pair. The empirical material requires two extensions to the current formulation. Fit is co-dependent: it depends on the alignment between tool capabilities and task requirements, and also on the user's instruction skills and routines. Two users working on the same task will produce different levels of fit if one has invested in upstream definition skill and the other has not. Task-Technology Fit also operates at multiple layers: at the individual task level can be high while at the process level is low, as demonstrated by the bottleneck identified in this study. Both extensions are necessary for AI generative tools, where the tool's behavior is dependent on the quality of the input and where the tool changes production rates without affecting the capacity of stages after. Both extensions are absent from the original formulation and are presented here as a refinement of the framework within the context of generative AI.

The third contribution is empirical grounding for the productivity paradox debate at the operational level. Brynjolfsson, Rock and Syverson (2019, pp. 1-2) argues that the gap between AI capability and all productivity is transferred by additional investments organizations have not yet made. Toma (2025) explains the same gap as a problem about

AI investment, and that those are running ahead before evidence of return. This study contributes with a concrete operational case in which the major complementary investments, such as structured training, governance norms, measurement infrastructure, and data readiness, are either absent or under-developed, and in which the value accumulates in intangibles that conventional accounting does not capture, in line with Corrado, Haskel, and Jona-Lasinio (2021, pp. 435-458). The case-level evidence connects an abstract macro-level debate to an identifiable set of organizational practices, and it does so for the operational level specifically, where existing AI research has been thin and has tended to rely on individual developer studies in controlled settings (Ziegler et al., 2022).

8.4 Practical Recommendations

The Practical recommendation will be directed at three groups of stakeholders, organized under the two research questions. Each recommendation will connect back to the findings and interpretations in the analysis and is intended to be actionable rather than aspirational.

8.4.1 Recommendations Addressing the Main RQ (Behavior and Operations)

The authors main recommendation would be to address the quality assurance and code review bottleneck. The improvements shown in individual efficiency are currently being absorbed by this bottleneck, and without alleviating the review-stage capacity, the organization have the possibly of not producing additional throughput. The authors believe that the automated testing solution which was brought up in the results is not the only relief mechanism. But, a combination of additional review capacity, a possible redesign of the procedures, and a selective automation of "lower-risk" portions of the review might suffice.

For Agile coaches and team leaders, the recommendation is to make the AI tool a standing item on the program for team forums and evaluation, so that the perceived gap of information between developer practice and Agile coaches does not persist. The agile coaches were not negligent; they were uninformed because no organizational mechanism was carrying the information up to them.

For BA Markets executive management, the recommendation is to commit to structured training programs and organizational learning routines, rather than relying on bottom-up practitioner knowledge.

8.4.2 Recommendations Addressing the Sub-RQ (Assessability of Financial Implications)

For the Operations management, the recommendation would lie in building basic measurement infrastructure to detect what passes through the pipeline. Cycle time from code to deployment, code review depth/length, change-failure rates, and possibly a "time to productive contributions" metric for new hires. The main issue seen is the overall absence of a measurement infrastructure.

For BA Markets executive management and CFOs in similar organizational contexts, the recommendation has two parts. The light-touch governance posture is defensible as an option-preservation strategy under genuine uncertainty about how technology will evolve. It is defensible only, however, when paired with the monitoring infrastructure required to exercise the option at value-maximizing moments. Holding an option without the instruments to know when to exercise is a half-completed strategy. A further recommendation is to invest in data readiness which the empirical material identifies as the ceiling that will bound further value captured once the QA bottleneck is relieved.

Evaluate AI tools in multi-level rather than single-level terms. Single-level evaluation will systematically miss either the gains, if conducted only at the aggregated level, or the costs, if conducted only at the individual level. The level problem identified in this study implies that the appropriate evaluation framework for generative AI tools must be designed to capture effects at, and between, organizational layers.

8.5 Societal and Ethical Implications

Three implications of the findings reach beyond the studied organization and are relevant to the wider societal debate on the development of generative AI in commercial settings.

The first concern the authors found was a displacement within labor and skill. The shift in cognitive work goes from doing to directing in the analysis; this has become a completely new skill and not just a simple productivity gain. The tasks that AI coding tools are most capable and reliably in doing have historically been the tasks that have given junior developers their entry into the profession. The skills that are needed to direct an AI are hard to learn because they require the exact entry-level experience that the AI is now replacing. The meaning of this is that the obvious productivity gain at the team level might come at the expense of the pipeline that is forming the next generation of developers. Acemoglu & Restrepo (2018, pp. 1-2) explain this as a central policy question raised by AI-driven automation. The case that is examined here illustrates this in a concrete matter. The point here is not that technology advancement should be resisted, but more the responsibility for sustaining the entry path into the profession and that these do not vanish only because individual productivity increases. The opposite fact is shown in the empirical material, that the newly onboarded developers are reaching higher productivity levels faster than prior to the implementation, which also in some way offset the concern at the entry in the profession. However, it does not address what might happen in the long run of professional maturity.

The second implication found concerns governance and accountability within critical infrastructure. Vattenfall is a major European energy company, and Vattenfall BA Markets operates in a trading environment where data accuracy and process reliability are crucial and have direct financial consequences. This study shows that GitHub Copilots adoption has travelled below the visibility line of the Agile Coaches (Block 2), that the AI coding tool is being integrated and used in operational systems without clear and organized governance. Also, the responsibility for those performing quality assurance and code review on AI assisted output is placed on individual developers and their own skillset rather than on documented procedures. None of these are seen as ethical failures, each one creates and represents a category of risk that becomes greater as the adoption scales and as the regulatory environment is moving towards stricter accountability for AI use in operational systems (EU AI Act). The recommendation the authors have is that the governance and documented procedures should grow alongside the adoption, rather than after a failure makes this gap visible.

The third implication concerns environment and resource costs. These AI assisted tools consume a substantial amount of energy and are reliant on a large IT infrastructure. The energy consumption association with the large IT infrastructure is itself a cost that cannot be ignored. The implication is not that Vattenfall BA Markets should reverse its adoption decision. Moreover, organization within the energy sector has specific reasons to consider the resource intensity of the AI tool they have integrated and taking that consideration into their own broader view of sustainability. The argument is more directed to the internal consistency between an organization's commitment to sustainability and its technology advancements decisions is a stakeholder concern, in particular for organizations that depend on their credibility in those commitments.

8.6 Limitation of the Study

The study has limitations that limit the conclusion to what it can support. They are presented here in particular rather than in general.

The study design is one company. The results are reliant on contextual factors of Vattenfall BA Markets such as a regulated energy sector, multinational scale, trading orientation, and maturity of the adjacent software organization. The relevant criterion of this study is transferability, not statistical generalizability. The study does not make any claim that the exact pattern discovered will replicate identically anywhere else. The issue discovered in this study is presented as a phenomenon that other organizations should investigate in their own organizational setting rather than as a universal rule that comes from one instance.

The data reflects a specific stage in the development of the company's adoption. P1's compounding gain track shows that the curve was still in motion at the time of our interviews. A study six months earlier or later would probably have given a different result of productivity improvements and perhaps a different evaluation of the severity of the bottleneck. The results are time stamped, and the authors highlight this as a strength feature of the design instead of a flaw.

The financial side of the study is more indirect instead of being showed straight away. Vattenfall BA Markets lacked operational monitoring that would allow the authors to perform a quantitative analysis of the flow through or of the bottleneck itself. This restriction comes from the measurement state of the researched organization; a different sub-question could be generated if the organization had a developed AI-tool monitoring system. The diagnostic framework that supports the sub-question, which is generalizable outside this limitation, provides more value than the particular, not yet assessable conclusion.

Finally, the sample of participants was restricted to 10 people inside one business sector. The sample was chosen for relevance and depth of knowledge, not for statistical coverage. This is consistent with the qualitative case-study design, but the lack of voices from outside the operations department, such as group level finance, IT security, or risk management, restricts the width the study can present on how the implementation appears from other vantage points.

8.7 Suggestions for Future Research

Empirically, a longitudinal study of the same organization over twelve to twenty-four months would capture the overall adoption curve as it continues to develop and would test whether the bottleneck identified in this study is resolved, persists, or migrates to a different stage of the pipeline. A multi-organizational study, drawing on cases outside of the energy sector, would test whether the level problem identified replicates in environments with different regulatory profiles, different operational tempos, and different company cultures. A focused empirical study of the quality assurance and code review stage in pipelines that have introduced AI coding tools would clarify whether the bottleneck is best relieved through additional reviewers, additional automation, redesign of the review process itself, or a combination of the three.

An additional empirical direction is the development and validation of measurable KPIs that can be linked to generative AI coding tools such as KPIs that capture code quality, an area in which the present study identified AI assistance as potentially impactful but did not quantify. KPI development for AI-coding tools is currently an open problem, and the absence of such instruments is the binding constraint identified by the sub-question of this study.

Theoretically, the multi-level extension of Task-Technology Fit proposed in this study can be developed further by formal modeling of the relationship between individual-level fit and process-level fit, and by comparative testing across cases. The integration of Real Options Theory with the measurement-infrastructure prerequisites for options exercise (underdeveloped in the existing option literature, where exercise is typically assumed to be costless) is a natural step, and the AI-implementation context is a productive setting for such work. The relationship between bounded-rationality satisficing at lower organizational levels and option-preservation logic at higher ones is a further theoretical thread that this case has surfaced but not exhausted.

Methodologically, mixed-method designs that combine qualitative data with operational assessments would directly extend the sub-question of this study. The measurement infrastructure that Vattenfall BA Markets has not yet built would be precisely the data layer needed for such designs. Future studies that gain access to such assessment, whether through partnerships with organizations that have invested in it, or through public data settings such as open-source software repositories, would be able to test the level problem quantitatively rather than describing the issue qualitatively. From the perspective of portfolio decision-making, P9 from this study makes the argument that this measurement can be applied outside of Vattenfall: Discussions over the ROI of AI tools are not possible.

Reference List

Acemoglu, D., & Restrepo, P. (2018). *Artificial intelligence, automation and work* (NBER Working Paper No.
24196). National Bureau of Economic Research. https://www.nber.org/system/files/working_papers/w24196/w24196.pdf

Agrawal, A., Gans, J., & Goldfarb, A. (2018). *Prediction machines: The simple economics of artificial intelligence*. Harvard Business Review Press.

Amazon Web Services. (2024). *What is generative AI?* Amazon Web Services. https://aws.amazon.com/what-is/generative-ai/

Autor, D. H. (2015). Why are there still so many jobs? The history and future of workplace automation. *Journal of Economic Perspectives*, *29*(3), 3–30.

Babashahi, L., Barbosa, C. E., Lima, Y., Lyra, A., Salazar, H., Argôlo, M., Almeida, M. A. de, & Souza, J. M. de. (2024, June 16). AI in the workplace: A systematic review of skill transformation in the industry. *MDPI*. https://www.mdpi.com/2076-3387/14/6/127

Barberis, N., & Thaler, R. (2002, September 19). A survey of Behavioral Finance. NBER. https://www.nber.org/papers/w9222

Benaroch, M., & Kauffman, R. J. (1999). A case for using real options pricing analysis to evaluate information technology project investments. *Information Systems Research*, *10*(1), 70–86.

Bell, E., Bryman, A., & Harley, B. (2022). Business research methods. Oxford University Press.

Braun, V., & Clarke, V. (2006, January). Using thematic analysis in psychology. https://www.researchgate.net/publication/235356393_Using_thematic_analysis_in_psychology

Bresnahan, T., & Trajtenberg, M. (1995). General purpose technologies: "Engines of growth"? *Journal of Econometrics*, *65*(1), 83–108.

Brinkmann, S., & Kvale, S. (2015). *Interviews: Learnings the craft of qualitative research interviewing* (3rd ed.). SAGE

Brynjolfsson, E., & McElheran, K. (2017). The rapid adoption of data-driven decision-making. *American Economic Review*, *106*(5), 133–139.

Brynjolfsson, E., Rock, D., & Syverson, C. (2019). *Artificial intelligence and the modern productivity paradox: A clash of expectations and statistics* (NBER Working Paper No. 24001). National Bureau of Economic Research. https://www.nber.org/system/files/working_papers/w24001/w24001.pdf

Caballar, R. D., & Stryker, C. (2026, April 13). *Pull request*. Ibm.com. https://www.ibm.com/think/topics/pull-request

Chlouverakis, K. (2024, April 26). *How artificial intelligence is reshaping the financial services industry*. EY. https://www.ey.com/en_gr/insights/financial-services/how-artificial-intelligence-is-reshaping-the-financial

Collis, J., & Hussey, R. (2021). *Business research: A practical guide for undergraduate and postgraduate students* (5th ed.). Bloomsbury Publishing.

Corrado, C., Haskel, J., & Jona-Lasinio, C. (2021). Artificial intelligence and productivity: An intangible assets approach. *Oxford Review of Economic Policy*, *37*(3), 435–458. https://doi.org/10.1093/oxrep/grab031

Davenport, T. H., & Ronanki, R. (2018). Artificial intelligence for the real world. *Harvard Business Review*, *96*(1), 108–116.

Dell'Acqua, F., McFowland, E., Mollick, E., Lifshitz, H., C. Kellogg, K., Rajendran, S., Krayer, L., Candelon, F., & R. Lakhan, K. (2026, January). This article was downloaded by: [199.94.42.12] on: 17 March 2026, at: 08:00. https://www.hbs.edu/ris/Publication%20Files/dell-acqua-et-al-2026-navigating-the-jagged-technological-frontier_5c589c8c-fbb5-458f-b285-c944746cd717.pdf

DeLone, W. H., & McLean, E. R. (2003). The DeLone and McLean model of information systems success: A ten-year update. *Journal of Management Information Systems*, *19*(4), 9–30.

Dixit, A. K., & Pindyck, R. S. (1994). *Investment under uncertainty*. Princeton University Press.

Dwivedi, Y. K., Kshetri, N., & Hughes, L. (2023, March). "so what if chatgpt wrote it?" multidisciplinary perspectives on opportunities, challenges and implications of Generative Conversational AI for Research, practice and policy. https://www.researchgate.net/publication/369170284_So_what_if_ChatGPT_wrote_it_Multidisciplinary_perspectives_on_opportunities_challenges_and_implications_of_generative_conversational_AI_for_research_practice_and_policy

Doody, O., & Noonan, M. (2013, May). Preparing and conducting interviews to collect data. https://pubmed.ncbi.nlm.nih.gov/23687846/

Dunwoodie, K., Macaulay, L. and Newman, A. (2023) 'Qualitative interviewing in the field of work and organisational psychology: Benefits, challenges and guidelines for researchers and reviewers', Applied Psychology: An International Review, 72(2), pp. 863–889. doi:10.1111/apps.12414. [Accessed 23 March 2026]

Dwivedi, Y. K. (2023, August). Opinion paper: "so what if chatgpt wrote it? ... https://www.sciencedirect.com/science/article/pii/S0268401223000233

Flyvberj, B. (2006, April). Five Misunderstandings About Case-Study Research. https://journals.sagepub.com/doi/10.1177/1077800405284363

Fonseca Costa, D., Melo Carvalho, F. de, Melo Moreira, B. C. de, & Santos Silva, W. (2020, July). *Confirmation bias in managerial decision-making: an experimental study with managers and accountants. View of confirmation bias in managerial decision-making: An experimental study with managers and accountants*. https://revistas.usp.br/rco/en/article/view/164200/169820

Ghoddusi, H., Creamer, G. G., & Rafizadeh, N. (2019). Machine learning in energy economics and finance: A review. *Energy Economics*, *81*, 709–727.

GitHub. (2025). *GitHub Copilot: Your AI pair programmer*. GitHub. https://github.com/features/copilot

Goodhue, D. L., & Thompson, R. L. (1995). Task-technology fit and individual performance. *MIS Quarterly*, *19*(2), 213–236.

Hagos, D., Battle, R., Rawat, D., & Hagos, D. (2024). *Recent advances in generative AI and large language models: Current status, challenges, and perspectives* [Preprint]. arXiv. https://arxiv.org/pdf/2407.14962

Hammersley, M. (2012). *What is qualitative research?* Bloomsbury Academic.

IBM. (2023, November 2). *What are large language models (LLMs)?* IBM. https://www.ibm.com/think/topics/large-language-models

International Energy Agency. (2022). *Gas market report, Q3-2022*. https://www.iea.org/reports/gas-market-report-q3-2022

International Energy Agency. (2025a). *What drives natural gas price volatility in Europe and beyond?* https://www.iea.org/commentaries/what-drives-natural-gas-price-volatility-in-europe-and-beyond

International Energy Agency. (2025b). *Gas market report, Q3-2025*. https://www.iea.org/reports/gas-market-report-q3-2025

Jiang, R., Nuzzo, R., & Roth, T. (2023). Unveiling the influence of artificial intelligence and machine learning on financial markets. Journal of Risk and Financial Management, 16(10), 434. https://www.mdpi.com/1911-8074/16/10/434

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263–291.

Kallio, H., Pietilä, A-M., Johnson, M., & Kangasniemi, M. (2016). Systematic methodological review: developing a framework for a qualitative semi-structured interview guide. Journal of advanced Nursing, 72(12), 2954-2965

Kandaurova, M. (2025). *Pursuing value creation through low-code AI* [Report]. Chalmers University of Technology. https://research.chalmers.se/publication/545423/file/545423_Fulltext.pdf

Kappes, A., Harvey, A. H., Lohrenz, T., Montague, P. R., & Sharot, T. (2019, December 16). *Confirmation bias in the utilization of others' opinion strength*. Nature News. https://www.nature.com/articles/s41593-019-0549-2

L. Mosier, K., & J Skitka, L. (1996, January). human decision makers and automated decision AIDS: Made for each other? https://www.researchgate.net/publication/230601064_Human_Decision_Makers_and_Automated_Decision_Aids_Made_for_Each_Other

Li, C. H., Boh, W. F., Ma, L., & Tang, P. M. (2025). Examining the effect of a firm's AI specialization on the technology firms it acquires: A real options perspective. *Journal of Management Studies*, *62*(2), 631–640.

Lincoln, Y. S., & Guba, E. G. (1985). naturalistic inquiry. Establishing Trustworthiness. https://ethnographyworkshop.wordpress.com/wp-content/uploads/2014/11/lincoln-guba-1985-establishing-trustworthiness-naturalistic-inquiry.pdf

Mayer, H., Yee, L., Chui, M., & Roberts, R. (2025, January 28). *Superagency in the workplace: Empowering people to unlock AI's full potential*. McKinsey & Company. https://www.mckinsey.com/capabilities/tech-and

McWilliams, A., & SIEGEL, D. (2012, March). Ent000176 submitted: March 28, 2012 © Academy of Management Journal. https://www.nrc.gov/docs/ML1208/ML12088A328.pdf

McWilliams, B., & Zachmann, G. (2026, March 4). *How will the Iran conflict hit European energy markets?* Bruegel. https://www.bruegel.org/first-glance/how-will-iran-conflict-hit-european-energy-markets

Microsoft. (2024, March 29). *Microsoft Copilot for Microsoft 365 overview*. Microsoft. https://learn.microsoft.com/en-us/copilot/microsoft-365/microsoft-365-copilot-overview

Nii Laryeafio, M., & Courage Ogbewe, O. (2023, August). Ethical consideration dilemma: Systematic review of ethics in qualitative data collection through interviews | journal of ethics in entrepreneurship and technology | emerald publishing. https://www.emerald.com/jeet/article/3/2/94/195210/Ethical-consideration-dilemma

Myers, S. C. (1977). Determinants of corporate borrowing. *Journal of Financial Economics*, *5*(2), 147–175.

Negrea, D. (2025, January 29). *Key statistics driving AI adoption in 2024*. HyperSense Blog. https://hypersense-software.com/blog/2025/01/29/key-statistics-driving-ai-adoption-in-2024/

Parasuraman, R., & Riley, V. (1997, June). Humans and automation: Use, misuse, disuse, abuse - Raja Parasuraman, Victor Riley, 1997. https://journals.sagepub.com/doi/10.1518/001872097778543886

Parker, C., Scott, S., & Geddes, A. (2019). Snowball Sampling. https://eprints.glos.ac.uk/6781/1/6781%20Parker%20and%20Scott%20(2019)%20Snowball%20Sampling_Peer%20reviewed%20pre-copy%20edited%20version.pdf

Peng, S., Kalliamvakou, E., Cihon, P., & Demirer, M. (2023). *The impact of AI on developer productivity: Evidence from GitHub Copilot* [Preprint]. arXiv. https://arxiv.org/abs/2302.06590

Precedence Research. (2025, November 20). *AI in energy market size to hit USD 75.53 billion by 2034*. Precedence Research. https://www.precedenceresearch.com/ai-in-energy-market

Queirós, A., Faria, D., & Almeida, F. (2017). Strengths and limitations of qualitative and quantitative research methods. *European Journal of Education Studies*, *3*, 369–387.

Queirós, A., Faria, D., & Almeida, F. (2017, September). strengths and limitations of qualitative and quantitative research methods. https://www.researchgate.net/publication/319852576_Strengths_and_Limitations_of_Qualitative_and_Quantitative_Research_Methods

Ransbotham, S., Khodabandeh, S., Kiron, D., Candelon, F., Chu, M., & Lafountain, B. (2020). *Expanding AI's impact with organizational learning* [Report]. MIT Sloan Management Review & Boston Consulting Group. https://web-assets.bcg.com/1e/4f/925e66794465ad89953ff604b656/mit-bcg-expanding-ai-impact-with

Regulation - 2016/679 - en - GDPR - EUR-lex. EUR. (2016, May 4). https://eur-lex.europa.eu/legal

Rubin, H. J., & Rubin, I. S. (2012). *Qualitative interviewing: The art of hearing data* (3rd ed.) SAGE

Saunders, M., Lewis, P., & Thornhill, A. (2009). *Research methods for business students* (5th ed.). Pearson Education.

Saunila, M., & Ukko, J. (2023, February). Value creation paths of organizations undergoing digital transformation - holopainen - 2023 - knowledge and process management - wiley online library. https://onlinelibrary.wiley.com/doi/full/10.1002/kpm.1745

Scharp, K.M. and Sanders, M.L. 2019. What is a theme? Teaching thematic analysis in qualitative communication research methods. Communication teacher. 33(2), pp.117–121.

Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, *69*(1), 99–118.

Spataro, J. (2023, March 16). *Introducing Microsoft 365 Copilot – your copilot for work*. Microsoft Blog. https://blogs.microsoft.com/blog/2023/03/16/introducing-microsoft-365-copilot-your-copilot-for-work/

Sukharevsky, A., West, A., Catania, C., & Grande, D. (2025, November 3). *How finance teams are putting AI to work today*. McKinsey & Company. https://www.mckinsey.com/capabilities/strategy-and-corporate-finance/our

Subramani, M., & Walden, E. (2001, May). The Impact of E-Commerce Announcements on the Market Value of Firms. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=914525

Toma, A. (2025). The impact of Artificial Intelligence on job roles and innovation. Theoretical and Applied Economics, 32(Special Issue), 161–166.

Tully, T. (2024, November 20). *2024: The state of generative AI in the enterprise*. Menlo Ventures. https://menlovc.com/2024-the-state-of-generative-ai-in-the-enterprise/

Vattenfall. (n.d.-a). *About us*. https://group.vattenfall.com/about-us

Vattenfall. (n.d.-b). *Our markets*. https://group.vattenfall.com/our-operations/our-markets/

Vattenfall. (n.d.-c). *Innovation – enabling fossil freedom*. https://group.vattenfall.com/our-operations/innovation/

Vattenfall. (2025, February 24). *How AI and robots are accelerating the energy transition*. https://group.vattenfall.com/press-and-media/newsroom/2025/how-ai-and-robots-are-accelerating-the-energy-transition/

Vattenfall. (2026). *Internal Reference, name not applicable.*

Vial, G. (2019, June). Understanding digital transformation: A review and a research agenda. https://www.sciencedirect.com/science/article/abs/pii/S0963868717302196?via%3Dihub

Woiceshyn, J., & Daellenbach, U. (2018, June). Evaluating inductive vs Deductive Research in Management Studies . https://www.emerald.com/qrom/article-abstract/13/2/183/361404/Evaluating-inductive-vs-deductive-research-in?redirectedFrom=fulltext

Yazan, B. 2015. Three Approaches to Case Study Methods in Education: Yin, Merriam, and Stake. Qualitative report. 20(2), pp.134–152.

Yin, R. K., & Campbell, D. T. (2018). *Case study research and applications: Design and methods*. SAGE Publications, Inc.

Ziegler, A., Kalliamvakou, E., Li, A., Rice, A., Rifkin, D., Simister, S., Sittampalam, G., & Aftandilian, E. (2022, June). Productivity assessment of neural code completion. https://dl.acm.org/doi/10.1145/3520312.3534864?gh_jid=3690456

Appendix

Interview Guide: Developers and Quantitative Analysts

Section 1 - Background and Role

Q1
Could you briefly describe your role, your main responsibilities, and the types of tasks that make up the majority of your working day?

Q2
How much of your work is routine and repetitive versus novel and explorative?
Do you work primarily individually or as part of a team on most tasks?

Q3
How often does your work require you to make judgement calls about architecture, implementation choices, or problem-solving approaches?

Q4
Before GitHub Copilot was introduced, how would you describe the way you typically approached your coding and development work?

Q5
Were there particular types of tasks you found especially time-consuming or cognitively demanding?

Q6
What sources of support or reference did you rely on - documentation, colleagues, Stack Overflow?

Section 2 - Adoption and Day-to-Day Use

Q7.
Could you walk me through how you actually use GitHub Copilot in a typical working day? When do you turn to it, and for what types of tasks?

- Are there tasks for which you use it consistently, and others where you tend not to use it at all?
- Has your pattern of use changed since you first started using it?

Q8.
How would you assess the quality of the suggestions GitHub Copilot provides in your specific domain of work?
- Do you find the suggestions reliable for the kinds of problems you work on, or do they require significant correction?

- Are there situations where the tool's suggestions are not useful — or could potentially cause problems if accepted without review?

Section 3 - Impact on How You Work and Think

Q9
Has GitHub Copilot changed the way you approach problems when writing code? If so, in what ways?

- Do you find yourself exploring solutions you might not have considered before the tool was available?
- Has it influenced how much time you spend searching for answers externally — for example, in documentation or online forums?

Q10
Thinking about tasks that are particularly complex or technically demanding — has GitHub Copilot changed how you handle those?

- Does the tool help you stay focused or does it introduce any distractions?
- Are there situations where having a suggestion readily available has caused you to overlook an alternative approach that might have been better?

Q11
Has using GitHub Copilot affected the confidence you have in the code you produce?

- Do you perform more or less manual review and testing of your code since adopting the tool?
- Have there been instances where you accepted a suggestion that later turned out to be incorrect or suboptimal?

Q12
Do you feel that working with GitHub Copilot has had any effect on your own technical development or learning?

- Are there areas where you feel you have learnt more because of the tool, or conversely, areas where you may be engaging less deeply than before?
- Is there a concern among developers that reliance on the tool might reduce the development of deeper technical skills over time?

Section 4 - Operational Outcomes

Q13
From your perspective, has GitHub Copilot made you more productive in your work? Can you give a concrete example of where you felt it had a clear positive effect?

- Are there specific task types - for example building data pipelines, debugging - where the productivity gain is most noticeable?

Has the speed at which you can deliver a working piece of code or analysis changed since adopting the tool?

Q14
Has GitHub Copilot influenced the quality or reliability of the outputs you produce, for: instance, the tools, scripts, or models you develop for operational use?

- Has the error rate in your code changed, to your knowledge?
- Does GitHub Copilots suggestions help produce more consistent or standardized solutions across the team, or does it introduce variability?
- Have you encountered any cases where Copilot produced suggestions that could pose operational or compliance risks if used without modification?

Q15
The work done in the Operations function has direct consequences for the business area's trading and risk management activities. From your perspective, do you think your work — potentially supported by Copilot — has had any noticeable impact on downstream processes (e.g., data quality, timeliness of reports, incident frequency)?

- Can you describe a situation where the quality or speed of your work, potentially supported by Copilot, influenced a downstream operational or commercial outcome?
- Conversely, is there a situation where limitations or errors related to the tool's suggestions created problems downstream?

Section 5 - Team and Organizational Dimension

Q16
How do your colleagues in the team use GitHub Copilot - is adoption relatively uniform, or does it vary considerably between individuals?
- Are there differences between more experienced and more junior developers in how the tool is used or perceived?
- Is there a shared understanding within the team of when it is appropriate to use the tool and when to be cautious?

Q17
Has the introduction of GitHub Copilot changed the way the team collaborates or communicates when working on shared code or projects?
- Do you think Copilot has changed the kinds of discussions the team has during planning, design sessions, or code reviews
- Has it changed how knowledge is shared between team members?

Q18

From your perspective, how well does the organization currently support and govern the use of GitHub Copilot - for example, in terms of guidelines, security considerations, or feedback on usage?

- Are there policies or guardrails in place that shape how the tool can be used, particularly given the sensitivity of the data and systems you work with?

- Is there a channel through which developers can give feedback on the tool or raise concerns about its performance?

Section 6 - Strategic and Future Perspective

Q19

Thinking about the next two to three years - how do you expect your use of GitHub Copilot, or AI-assisted coding tools more broadly, to evolve within your work?

- Are there aspects of your current work that you believe could be further supported or automated through tools like this?
- Do you see this as an area where Vattenfall BA Markets could build a meaningful competitive advantage, or is it more a question of keeping pace with what everyone else will adopt?

Q20
If you were advising the organization on how to get more value from its investment in GitHub Copilot, what would your main recommendations be?

- Are there barriers currently that prevent the tool from being used as effectively as it could be?
- What would need to change - in terms of support, training, or governance - for the tool to deliver more value?

Q21
Do you foresee any skills or competencies becoming more or less important because of continued GitHub Copilot use?

Section 7 – Closing questions

Q20
We have now covered the main areas we hoped to explore. Before we finish, we have two final questions.

Q21
Is there anything related to your experience with GitHub Copilot - or with the broader question of how AI tools are reshaping work in the Operations function - that we have not touched on, but that you think is important for us to understand?

Interview Guide: Agile Coaches

Section 1 - Background and Role

Q1
Could you describe your role as an agile coach, your main responsibilities, and how you typically interact with development teams?

- How many teams do you work with, and in what capacity (coaching, facilitation, advisory)?
- To what extent is your work focused on team processes versus technical practices?
- How involved are you in shaping how teams make decisions about tools, ways of working, or quality standards?

Q2
Before GitHub Copilot was introduced, how would you describe the way development teams
typically approached coding and problem-solving?

Were there recurring bottlenecks or decision challenges you observed (e.g. speed vs quality,
rework, reliance on key individuals)?

How did teams typically seek support or knowledge when facing unfamiliar technical problems?

Section 2 - Adoption and Day-to-Day Use

Q3
From your perspective as an agile coach, how has GitHub Copilot been adopted across the
teams you work with?

- Was adoption driven bottom-up by developers, top-down by management, or a mix?
- Have you observed differences in adoption between teams or roles?

Q4
How do you see GitHub Copilot being used in day-to-day development work at a team level?
Are there specific types of tasks or phases (e.g. implementation, refactoring, prototyping)

- where its use seems most common?
- Are there situations where teams deliberately avoid using the tool?

Section 3 - Impact on Decision-Making and Ways of Working

Q5
Have you observed any changes in how teams make technical or implementation-related decisions since GitHub Copilot was introduced?

- For example, changes in how quickly decisions are made, or how many alternatives are considered?
- Do teams appear more or less reliant on discussion and peer review?

Q6
From your observations, has GitHub Copilot influenced how teams deal with complexity or
uncertainty in development work?

- Does it appear to reduce cognitive load, or does it sometimes introduce new risks or shortcuts?
- Have you seen cases where readily available suggestions influenced teams to converge too quickly on a solution?

Q7
How has the introduction of GitHub Copilot affected practices around review, testing, or quality assurance from a team perspective?

- Have expectations around code review changed?
- Do teams compensate for AI assistance in any explicit way?

Section 4 - Operational Outcomes

Q8
From your perspective, has GitHub Copilot had any noticeable impact on team productivity or delivery flow?

- For example, changes in cycle time, throughput, or perceived efficiency
- Are there particular types of work where the impact is most visible?

Q9
How has GitHub Copilot influenced the quality, robustness, or maintainability of the solutions teams deliver?

- Have you observed changes in defect patterns, rework, or technical debt discussions?
- Do teams appear more aligned or more divergent in how solutions are implemented?

Q10
The teams you support work in a context with strong operational and risk implications. Have
you seen any downstream operational effects related to AI-assisted coding?

- For example incidents, compliance concerns, or improvements in data quality or reliability
- How are such risks identified and handled at a team or organizational level?

Section 5 - Team and Organizational Dimension

Q11

How has GitHub Copilot affected collaboration and knowledge sharing within or between teams?

- Has it changed what is discussed in refinements, retrospectives, or technical forums?
- Does it alter the balance between individual expertise and shared team knowledge?

Q12
From your perspective, how mature is the organization's governance around GitHub Copilot
today?

- Are there clear norms, principles, or guardrails for its use?
- How are ethical, security, or compliance considerations addressed in practice?

Q13
What role do agile coaches play - or should they play - in shaping how tools like GitHub Copilot are used responsibly and effectively?

- Is this currently part of your mandate, or more implicit?

Section 6 - Strategic and Future Perspective

Q14
Looking ahead at two to three years, how do you expect AI-assisted coding tools to influence
agile ways of working at Vattenfall BA Markets?

- Do you expect changes in team autonomy, skill requirements, or roles?
- Could this influence how work is planned or decomposed?

Q15
From a strategic perspective, do you see GitHub Copilot as creating new organizational options or capabilities?

- For example, faster experimentation, scaling development capacity, or reducing dependency on scarce skills
- Or is it mainly a tool for incremental efficiency?

Q16
If you were advising senior management, what would you recommend to maximize the value
of GitHub Copilot while managing its risks?

- In terms of coaching, training, governance, or measurement
- What would need to change for the organization to fully realize its potential?

Q17
Do you foresee any skills or competencies becoming more or less important for teams because of continued use of AI-assisted coding tools?

Section 7 – Closing question

Q18

Is there anything related to GitHub Copilot, team decision-making, or AI's influence on agile practices that we have not discussed but that you believe is important?

Interview Guide: Project Portfolio Manager

Section 1: Operational Visibility & Delivery Impact

Could you briefly describe your role, your main responsibilities, and the types of tasks that make up the majority of your working day?

How has the introduction of GitHub Copilot affected the project timelines within your portfolio? Is there any effect at all?

Q1.
From a portfolio perspective, if developer output has increased faster than downstream processes (review, testing, deployment), what does that constrain most today?

- Speed to market,
- Delivery predictability,
- Risk tolerance, or
- The number of initiatives you can safely run in parallel?

Q2.
Automated testing has been discussed as a potential solution, but the coaches noted significant barriers given the sensitivity of your systems - From a portfolio perspective, is this trade-off framed more as a risk decision or an investment decision?

Are there parts of the codebase or pipeline where automated testing is already viable, and others where it is not?

Section 2: Financial Evaluation & Business Case

Q3.
The implementation of GitHub Copilot and other AI tools should be seen as an ongoing investment rather than a one-time deployment — how does the organization or you currently evaluate whether to invest further, and who drives that decision?

Q4
One of our participants noted that productivity gains were modest at first but grew substantially over time as prompting skill developed. Does the organization's current evaluation framework account for that kind of compounding return, or is it more of a point-in-time assessment?

Q5.
Has the productivity increase at the individual level translated into any visible change at the output level of your function - for example, are you delivering more with the same team, or have resourcing decisions been influenced by the availability of GitHub Copilot?

If the gains are not yet visible at your level, where do you think they are being absorbed?

Has there been any instance where the tool has directly affected a budget or resourcing conversation?

Q6.
Is GitHub Copilot factored into capacity planning or headcount discussions — in terms of what a team of a given size can now deliver compared to before?

- If not, do you think it should be — and what would need to change for that to happen?

Are there any risks you see in factoring AI tool productivity into capacity assumptions at this stage?

Section 3: Evidence of Business Value

Q7.
Are there any examples you are aware of where GitHub Copilot created clear business value - whether through new capabilities, faster model deployment, or reduced project timelines?

With knowledge of the implications from task level usage.

- Were any of the examples described by our developer participants - such as the solar forecasting model built in an afternoon - something that reached your level, or did they stay within the team?
- Have trading or other commercial functions noticed or commented on changes in the speed or quality of outputs coming from your teams?

Q8.
If this value largely stays invisible, do you see that as affecting the quality of portfolio decisions – for example in prioritization, funding, or risk assessment?

- If it stays invisible, is that a gap you see as worth addressing - and if so, how?
- Are there decisions that you might have made differently if that value had been more visible to you earlier?

Section 4: Decision-Making & Strategic Scope

Q9.
Has GitHub Copilot changed how you think about what your teams can take on - in terms of project scope, team size needed, or timelines you are willing to commit to?

Can you give a concrete example of a project or initiative where your expectations or commitments were shaped by the availability of AI tools?

- Has it changed how you evaluate build-versus-buy decisions for new capabilities?

Q10.
Has the implementation GitHub Copilot changed how you think about optionality in the portfolio – for example, starting initiatives earlier, running mor experiments, or deferring irreversible commitments?

Q11.

When decisions are made faster at the deployment level, where do the financial effects show up first – costs, revenue, or risk exposure?

Interview Guide: Operations Managers
Director of Analysis, and Director Business Infrastructure

Section 1: Operational Visibility & Delivery Impact

Q1.
From our talk with the agile coaches last week, they described a bottleneck forming at the QA and code review stage - is that something that is occurring across the department, or only in the agile teams?

Do you see this as a resourcing problem, a process problem, or a structural consequence of AI-accelerated code production?

Has this been flagged to you previously, or is this the first time you have heard about it at this level of detail?

Q2.
Is there a strategy in place, or being considered, for scaling review and testing capacity to match the increased output from developers using AI tools?
- Who within the organization would own that decision - is it something that sits with you, or does it require escalation?
- Have teams been given any guidance on how to self-manage the increased review load in the meantime?

Q3.
Automated testing has been discussed as a potential solution, but the coaches noted significant barriers given the sensitivity of your systems - what are your thoughts on that tradeoff?
- Are there parts of the codebase or pipeline where automated testing is already viable, and others where it is not?
- Is the reluctance around automation primarily a technical constraint, a security constraint, or a cultural one?

Section 2: Financial Evaluation & Business Case

Q4.
The implementation of GitHub Copilot and other AI tools should be seen as an ongoing investment rather than a one-time deployment - how does the organization currently evaluate whether to invest further, and who drives that decision?

One of our participants noted that productivity gains were modest at first but grew substantially over time as prompting skill developed. Does the organization's current evaluation framework account for that kind of compounding return, or is it more of a point-in-time assessment?
- More tokens?
- More licenses?

Q5.
Has the productivity increase at the individual level translated into any visible change at the output level of your function - for example, are you delivering more with the same team, or have resourcing decisions been influenced by the availability of GitHub Copilot?

- If the gains are not yet visible at your level, where do you think they are being absorbed?
- Has there been any instance where the tool has directly affected a budget or resourcing conversation?

Q6.
Is GitHub Copilot factored into capacity planning or headcount discussions - in terms of what a team of a given size can now deliver compared to before?

- If not, do you think it should be - and what would need to change for that to happen?
- Are there any risks you see in factoring AI tool productivity into capacity assumptions at this stage?

Section 3: Evidence of Business Value

Q7.
Are there any examples you are aware of where GitHub Copilot created clear business value - whether through new capabilities, faster model deployment, or reduced project timelines?

With knowledge of the implication from task level usage. Were any of the examples described by our developer participants – such as the solar forecasting model built in an afternoon - something that reached your level, or did they stay within the team?

- Have trading or other commercial functions noticed or commented on changes in the speed or quality of outputs coming from your teams?

Q8.
More broadly, how does value created at the developer level get communicated upward — is there a mechanism in place for that, or does it largely stay invisible at your level?

- If it stays invisible, is that a gap you see as worth addressing - and if so, how?
- Are there decisions that you might have made differently if that value had been more visible to you earlier?

Section 4: Decision-Making & Strategic Scope

Q9.
Has GitHub Copilot changed how you think about what your teams can take on - in terms of project scope, team size needed, or timelines you are willing to commit to?

- Can you give a concrete example of a project or initiative where your expectations or commitments were shaped by the availability of AI tools?
- Has it changed how you evaluate build-versus-buy decisions for new capabilities?

Controlling & Strategic Development

Section 1: Financial Visibility of the Tool

Q1.
How is the cost of GitHub Copilot currently accounted for within BA Markets - is it a centralized IT cost, a dedicated budget line, or distributed across teams?

- Does that cost structure give you visibility into how intensively the tool is being used, or is it a flat license cost regardless of usage?
- Has the cost grown since initial deployment - for example through premium tier expansions - and if so, how has that been managed?

Q2.
Is the financial impact of GitHub Copilot something you actively monitor, or does it currently sit below the threshold of formal financial scrutiny?

- If it is not actively monitored, what would need to change for it to be?
- Does the absence of formal financial tracking affect how investment decisions around the tool are made?

Section 2: Investment Evaluation & Decision-Making

Q3.
The implementation of GitHub Copilot should be seen as an ongoing investment rather than a one-time deployment - how does the organization currently evaluate whether to invest further, and who drives that decision?

- Is that evaluation based on qualitative signals, or are there quantitative thresholds that need to be met to justify further spend?

One of our developer participants described productivity gains as modest at first but growing substantially over time as prompting skill developed - does the current evaluation framework account for that kind of compounding return, or is it a point-in-time assessment

Q4.
What would a credible business case for increasing investment in GitHub Copilot need to include — what evidence or metrics would you want to see?

Is there a standard investment evaluation methodology used within BA Markets that would apply here, for example, payback period, NPV, or cost-per-output metrics?

- Who would be responsible for constructing that business case - is it a finance responsibility, a technology responsibility, or a shared one?

Q5.

Has the way the organization makes investment decisions around technology changed at all since AI tools became available - or is GitHub Copilot being evaluated through the same framework as any other tooling investment?

- If the framework has not changed, do you think it should - given that AI tools have a different productivity trajectory than conventional software?
- Is there a risk that a standard investment evaluation framework underestimates the value of a tool whose returns compound over time?

Section 3: Operational Changes with Financial Consequences

Q6.
If developers are producing significantly more output per unit of labor input, as our interviews suggest, how should that show up in the cost base - and is it currently showing up at all?

- Is there a mechanism for translating individual productivity gains into observable cost reductions, or do those gains tend to be absorbed into expanded scope rather than reduced cost?
- Have you seen any financial signal of that kind of impact in the numbers - or is it still too early or too diffuse to observe?

Q7.
The agile coaches described a bottleneck forming at the QA and code review stage, where code is being produced faster than it can be reviewed - does that kind of operational friction have a visible cost consequence?

- If delivery speed is not improving despite higher developer output, what does that mean for the financial return on the GitHub Copilot investment?

Would resolving that bottleneck (for example through investment in automated testing) be something that would be evaluated as a follow-on investment decision?

Section 4: Labour Cost & Resourcing Implications

Q8.
Has GitHub Copilot influenced any headcount or resourcing decisions within BA Markets - either in terms of hiring fewer people, redeploying existing staff, or expecting more from teams of the same size?

- If not, do you think it should be factored into those decisions - and what would need to be true for that to happen responsibly?
- Is there a risk that the organization is making resourcing assumptions based on AI productivity gains that may not yet be fully materializing?

Q9.
Is GitHub Copilot currently factored into capacity planning - in terms of what a team of a given size is expected to deliver?

- If productivity assumptions have shifted, are those shifts being made explicitly and formally, or are they implicit?

- Who holds accountability for ensuring that capacity assumptions reflect the actual - rather than theoretical - productivity impact of the tool?

Section 5: Financial Implications of Governance Gaps

Q10.
If the return on the GitHub Copilot investment depends heavily on how well individual developers use the tool, does the absence of structured training represent a financial risk - in the sense that the organization is paying for a capability it is not fully extracting?

- Is the cost of not training users being considered alongside the cost of the tool itself?
- Would a formal training program be evaluated as a cost item that enhances the return on the existing GitHub Copilot investment, or would it be treated as a separate expenditure?

Q11.
More broadly, how does the organization account for the cost of unrealized value - situations where a tool is in place but not being used to its full potential due to governance or capability gaps?

- Is there a mechanism for surfacing that kind of hidden cost to decision-makers?
- Does that framing - of governance gaps as a financial issue rather than just an operational one - resonate with how you think about the investment?

What costs are associated with the implementation of GitHub Copilot?
Are you affected by the operational implications described by developers and agile coaches?

- Bottle neck effect?
- Productivity?

Can you see indirect cost reduction or increase changes in development or operational decision-making?

- Higher productivity
- More pull requests
- Higher usage of GitHub copilot
- Is cost avoided rather than reduced?

Interview Guide: Executive management

Q1. Role and Decision Context

To begin with, could you describe your role within BA Markets and the types of decisions you are primarily responsible for?

Q2. Organizational Context Before AI-Assisted Tools

From a management perspective, what were the main constraints on delivery speed, quality, or flexibility?
How predictable were development outcomes in terms of timelines, capacity, and resource needs?

Q3. Observed Changes at the Organizational Level

Over the past year, have you observed any changes in:

- Delivery speed
- Volume of initiatives
- Expectations on development or analytics teams?

When such changes occur, how do they typically become visible to you (e.g., through escalations, planning cycles, budget discussions)?

To what extent do you associate these changes with new digital or AI-enabled ways of working, even if the connection is indirect?

Q4. Decision Behavior

Has the increased pace or flexibility of development changed how you make decisions related to:
- Prioritization
- Sequencing of initiatives
- Risk acceptance?

Compared to before, do today's decisions rely more on:

- Delivering something quickly and iterating?
- Or getting things "right" before resources are committed?

Have expectations on teams changed in terms of how quickly ideas should move from concept to implementation?

Q5. Operational Implications at the System Level

Have you observed any tension between faster development and downstream processes such as:

- Reviews
- Testing
- Approvals
- Regulatory compliance?
- If invisible to you: how would the organization respond?

When such imbalances arise, how does the organization typically respond?
- Adding capacity?
- Adjusting expectations?
- Accepting delays?

Do you see these effects as temporary growing pains, or as signals that existing structures need to change?

Q6. Financial Logic and Investment Rationale

From a financial perspective, how are AI-enabled development capabilities currently viewed?
- As an incremental cost?
- As infrastructure?
- As a strategic capability?

Are these types of investments evaluated using the same logic as traditional IT investments?
Why, or why not?

How do you think about "return" when the benefits are:
- Indirect
- Distributed
- Difficult to isolate?

If this investment were wound down or scaled back tomorrow, what would the organization lose, and how would you notice it?

Probing sub-questions:
- Would the loss be primarily financial, operational, or positional (i.e., falling behind competitors)?
- Are there switching costs or dependencies on capabilities that have already been built up?

Does the organization have visibility into what it has become dependent on, which was not anticipated at the time of investment?

Follow-up:
Would it be reasonable to expect precise ROI figures for such tools?

Q7. Portfolio and Strategy Implications

Has increased development flexibility changed how ambitious the organization can be with regard to:

- Number of initiatives
- Time-to-market
- Strategic optionality?

Do you believe that current portfolio and governance processes are designed for this kind of flexibility?

- If not, what tensions does that create at the senior leadership level?

Q8. Risk, Regulation and Accountability

From your perspective, does increased reliance on advanced tools change how risk is distributed within the organization?

How important is clear accountability when outcomes are partly shaped by automated or AI-enabled processes?

Do you see a need for stronger governance – or would that risk slowing down value creation?

Q9. Forward-Looking Reflections

Looking ahead, how do you expect AI-assisted tools to affect:

- Management decision-making
- Organizational design
- Investment priorities

What do you see as the greatest opportunity – and the greatest risk – if development continues to accelerate?

If you were to advise colleagues in similar organizations, what guidance would you offer regarding AI-enabled development capabilities?

Tables and figures

Each Theory Addresses a Different Layer

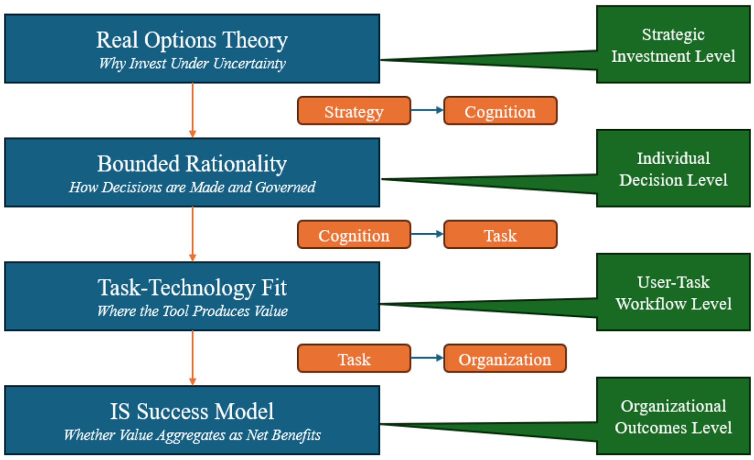

Figure 1

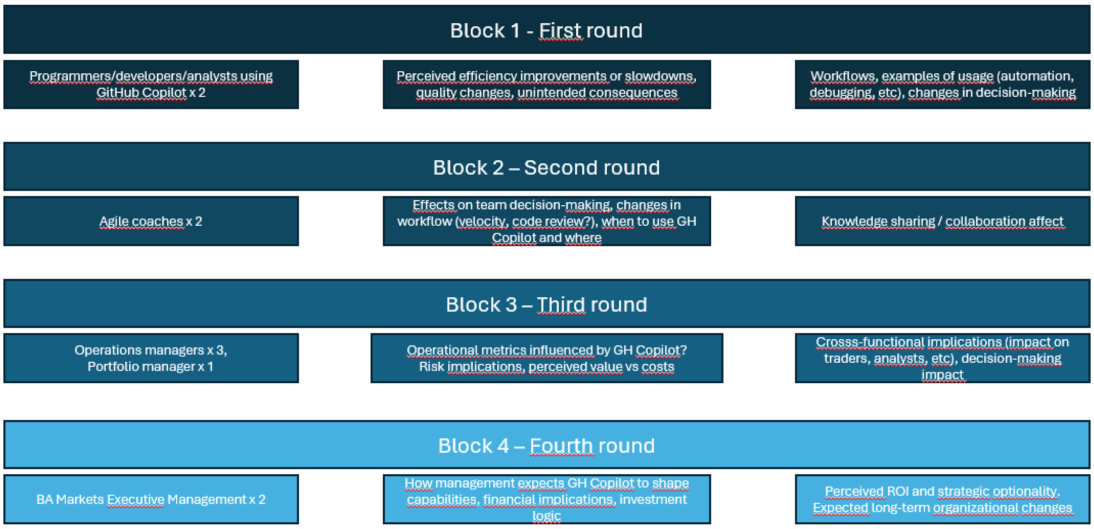
Block 1 - First round
Programmers/developers/analysts using GitHub Copilot x 2
Perceived efficiency improvements or slowdowns, quality changes, unintended consequences
Workflows, examples of usage (automation, debugging, etc), changes in decision-making
Block 2 – Second round
Agile coaches x 2
Effects on team decision-making, changes in workflow (velocity, code review?), when to use GH Copilot and where
Knowledge sharing / collaboration affect
Block 3 – Third round
Operations managers x 3, Portfolio manager x 1
Operational metrics influenced by GH Copilot? Risk implications, perceived value vs costs
Crosss-functional implications (impact on traders, analysts, etc), decision-making impact
Block 4 – Fourth round
BA Markets Executive Management x 2
How management expects GH Copilot to shape capabilities, financial implications, investment logic
Perceived ROI and strategic optionality. Expected long-term organizational changes

Figure 2

Table 1

Participant	Role	Block	Experience	Interview Time and Date
P1	Lead/Fundamental Analyst	Block 1: Programmer & Analyst	Senior	45-60 min 30-03-2026
P2	Developer (Application & Pipelines)	Block 1: Programmer & Analyst	Senior	45-60 min 31-04-2026
P3	Agile Coach	Block 2: Agile Coach	Senior	45-60 min 10-04-2026
P4	Agile Coach	Block 2: Agile Coach	Senior	45-60 min 10-04-2026
P5	Controlling & Strategic Development	Block 3: Management	Senior	45-60 min 14-04-2026
P6	Director of Analysis	Block 3: Management	Senior	45-60 min 14-04-2026
P7	Director Business Infrastructure	Block 3: Management	Senior	45-60 min 14-04-2026
P8	Project Portfolio Manager	Block 3: Management	Senior	45-60 min 15-04-2026
P9	Vice President Controlling & Strategy Development	Block 4: Executive Management	Senior	30-45 min 20-04-2026
P10	Vice President Operations	Block 4: Executive Management	Senior	30-45 min 23-04-2026

Table 2

Theme finding	RQ Component	Main Finding
1. Adoption & Individual fit	Sub RQ, Assessability	Substantial individual productivity gains: Approx 30%-40% on developer's tasks, with order of magnitude. reductions on specific applications
	Main RQ, Behavioral	Developers now spend more time defining what to build and less time writing the code itself.
	Main RQ, Behavioral	Developers learn the tool on their own, as there was no formal training, and views in organizational support diverge.
2. Workflow disruption & bottlenecks	Main RQ, Behavioral	Testing and code review have become the bottleneck in the delivery.
	Main RQ, Behavioral	Faster coding has not led to faster delivery.
	Main RQ, Behavioral	Agile coaches did not know everyone on their teams was using GitHub Copilot.
3. Decision-making under uncertainty	Sub RQ, Assessability	The investment is justified by belief, not measurement; no baseline or evaluation framework exists.
	Sub RQ, Assessability	Governance is deliberately light to encourage adoption, but there are no instruments to track results.
	Main RQ, Behavioral	Managers acknowledge the bottleneck but have not formally addressed it.
4. Financial & Operational implications	Sub RQ, Assessability	The license cost is too small to scrutinize; GitHub Copilot is treated as enablement, not a cost-saving tool.
	Sub RQ, Assessability	Productivity gain cannot be traced to financial or delivery outcomes but hiring and team composition are already shifting.
	Sub RQ, Assessability	Data quality and accessibility, not the tool itself, will limit further value
Cross block analysis		Effects do not add up across all block levels; individual gain, team workflows, and management decisions are disconnected.