A Work Project, presented as part of the requirements for the Award of a Master's degree in Finance, from the Nova School of Business and Economics.

# ROI Framework Development for AI Impact Assessment – Decision-Making Roles

Olga Lagereva (60121)

Work project carried out under the supervision of:

Milton de Sousa

12-2025

**Abstract**

This study aims to develop a ROI framework to assess the impact of AI tools implemented within organizations. Theoretical evidence provides insights into main approaches and variables necessary to design the methodology. Expert interviews allow for an evaluation and enhancement of the introduced approach. The resulting framework adopts a hybrid structure to capture financial, productivity, human-centered, and contextual metrics. This enables organizations to observe measurable impacts of AI, investing in technologies that accelerate revenue growth.

**Keywords:** Artificial Intelligence (AI), AI Tools, Return on Investment (ROI), Process Optimization, Workflow Automation, Human-AI Collaboration

This work was funded by Fundação para a Ciência e a Tecnologia (UID/00124/2025, UID/PRR/124/2025, Nova School of Business and Economics) and LISBOA2030 (DataLab2030 - LISBOA2030-FEDER-01314200).

**Table of Contents**

**Group Contribution:**

# 4 Theoretically Derived Framework (Pre-Expert Interviews)

## 4.1 Framework Design and Conceptual Structure

The first version of the framework is derived from theoretical evidence before being validated through interviews with experts. The research employs a logic-model structure that traces causal links from inputs to processes, outputs, outcomes, and finally ROI. Each layer defines measurable constructs and mechanisms through which effects propagate, allowing the framework to serve as a structured guide for future empirical measurement.

**Inputs** represent the resources and enablers required to operate agentic AI, including licenses, computing capacity, integration work, data preparation, and oversight requirements.

**Processes** capture the execution of work once AI is embedded, such as decision cycles, human-in-the-loop checkpoints, and autonomy ratios.

**Outputs** denote the immediate products of these processes, for example speed, quality, and number of analytical artefacts or decisions generated.

**Outcomes** reflect the broader organizational effects that emerge when AI-mediated outputs influence stakeholders. These include improved decision accuracy, coordination efficiency, user satisfaction, and capability development. The framework also accounts for spillover benefits, in which enhanced coordination or data quality positively affect colleagues and teams not directly using the AI agent.

Finally, **ROI** integrates these effects into monetized value through time savings, cost avoidance, and revenue or margin improvements, offset by investment and risk-related costs.

Applying this structure to agentic AI makes it possible to articulate not only direct operational effects but also second-order impacts such as process redesign, learning dynamics, and coordination improvements. By grounding the framework in established logic-model methodology, the study ensures that its causal assumptions are transparent, testable in future empirical work, and aligned with standard practices in evaluation science rather than based solely on conceptual intuition.

### 4.2 Cross-Cutting Adjustments

Four cross-cutting mechanisms increase analytical robustness. First, attribution is established conceptually through causal mapping and the identification of potential confounders, acknowledging that empirical isolation will occur only in later applications. Second, risk and quality adjustments ensure that expected costs of error remediation, compliance, or model drift are incorporated when evaluating projected ROI. Third, learning-curve effects distinguish short-term ramp-up costs from steady-state performance. Fourth, human-capital dynamics are explicitly recognized: automation is treated not merely as labor saving but as value reallocation, allowing employees to devote the time freed by AI to higher-quality or more strategic work such as creative analysis or innovation.

### 4.3 Measurement Logic and Indicator Design

The framework specifies how future implementations could operationalize each layer through representative indicators. Inputs would be measured through total cost of ownership, agent coverage, and data readiness; process measures could include cycle time, rework rate, and

autonomy ratio; outputs might capture throughput and first-pass quality; and outcomes could track decision accuracy, stakeholder satisfaction, and risk incidents. Although these indicators are not yet empirically populated, they form the measurement logic that underpins the framework's validity. The interviewees in the validation stage assess the relevance of these proposed indicators, providing qualitative evidence of construct adequacy.

### 4.4 Hybrid ROI Framework

To formalize how diverse effects can be integrated into a single valuation structure, the study adopts a hybrid ROI formulation frequently used in digital transformation and AI evaluation research (Legner et al. 2017). ROI models combine productivity-based and direct financial benefits to capture both tangible and intangible contributions to organizational performance. This is particularly suitable for agentic AI, where value creation often arises not only from measurable time savings but also from improvements in decision quality, coordination, error reduction, and knowledge flow. These benefits are directly linked to the interaction of employees with technology in the workplace. Therefore, the adoption rate helps to account for human context and measure the level of human-AI collaboration, transforming the formula into a hybrid approach.

The conceptual ROI equation applied is:

$$ROI_{AI} = \frac{\left( B_{prod} + B_{fin} \right) \times A_{rate} - C_{total}}{C_{total}} \quad (1)$$

where:

- $B_{prod}$ represents productivity-based benefits (e.g., reduced cycle times, time saved, lower rework),
- $B_{fin}$ represents direct financial benefits (e.g., cost avoidance, margin improvements, incremental revenue),
- $A_{rate}$ represents adoption rate (e.g., user satisfaction, trust, usefulness, cognitive abilities),
- $C_{total}$ denotes the total cost of ownership, including licensing, integration, oversight, and risk-management costs.

The next chapter presents the qualitative validation results, summarizing how practitioner feedback confirmed, questioned, or refined specific elements of the proposed framework. With this input, a refined version of the formula was created. The discussion then draws broader implications for future empirical testing once operational data becomes available.

# 6 Validated Framework (Post-Expert Interviews)

## 6.1 Framework Redesign and Indicator Enrichment

Based on expert validation, the proposed **list of indicators** proved to be clear and relevant to estimate the value of AI tools. Participants recommended its further expansion with three additional indicators. The first is opportunity cost which goes into ROI as part of financial benefits associated with AI adoption in organizations. For instance, it allows for the reallocation of specialists and resources to focus on innovative projects, thereby fostering research and development. The second factor includes throughput efficiency that measures the performance of AI agents, thus being part of processes. It can assess the contribution of AI to timely execution of projects or higher resolution rates. The third factor is time spent on process redesign before companies invest in AI, thus this goes into inputs. It involves the recruitment of consultants and analysts to revise and improve company workflows. This eventually translates into the cost of implementation in the ROI formula.

Another enhancement is classification of processes into **generative or classic categories**, using either reasoning or a given set of calculations. Examples of generative types include autonomy ratio, decision cycle time, escalation rate, and interpretive error tracking. This refers to AI agents that make independent decisions and interpret uncertainties, reporting them to humans. Examples of classic types include repetition rate, workflow bottlenecks, throughput efficiency, and operational error tracking. This represents routine tasks and observation of standard workflows to avoid mechanical issues. Any further modification of the indicator list should be made according to business type and company size.

Expert evaluation has also shown that the existing **ROI formula** is comprehensive but enhancing it with additional variables can capture a broader impact of AI. Participants suggest adding four key variables, organizational maturity, human engagement and motivation, quality improvement, and brand reputation. Prior to their inclusion in the formula, each should be carefully examined. The benefits in terms of brand reputation are hard to measure, especially since they are affected by multiple external factors other than AI. The changes in quality are indirectly measured through existing variables, carrying the risk of double counting. For instance, productivity benefits capture the reduction of errors and rework, implying higher quality. Next, a measure of human engagement and motivation is closely related to the adoption rate, making it reasonable to combine these variables into one. This would allow companies to observe whether employees working with AI tools are truly motivated to use them. Another variable proposed by the interview experts was organizational maturity, referring to the company's existing technological infrastructure (e.g., processing capacity, computational hardware). While we acknowledge that an organization's technological baseline is a major cost driver, we also recognized that this factor cannot be assessed in isolation. Instead, it must be considered in relation to the complexity of the AI solution to be implemented. An organization with limited technological capabilities may incur substantial costs when implementing a highly complex AI tool. Conversely, if the intended AI solution is relatively simple, the existing infrastructure may be sufficient and therefore not constitute a significant barrier. In such cases, the simplicity of the AI tool and the organization's lower level of technological maturity effectively offset one another. For this reason, we introduce a combined variable, the fit, which captures the alignment between organizational maturity and AI complexity.

Following this evaluation, the ROI framework transforms into:

$$ROI_{AI} = \frac{(B_{prod}+B_{fin}) \times A_{human} \times M_{fit} - C_{total}}{C_{total}} \quad (2)$$

where,

- $B_{prod}$ represents productivity-based benefits (e.g., time savings, throughput efficiency, compliance, defect and risk reductions),
- $B_{fin}$ represents direct financial benefits (e.g., cost avoidance, incremental revenue, opportunity cost),
- $A_{human}$ represents human adoption and motivation (e.g., user satisfaction and motivation, trust, non-autonomous usage, task replacement, job security),
- $M_{fit}$ represents the fit between the organization's technological maturity and complexity of the AI (e.g. processing capacity, computational load of AI)
- $C_{total}$ represents the total cost of ownership (e.g. integration and deployment, maintenance, licensing, process redesign).

The new approach accounts for contextual metrics, keeping a hybrid scope of financial, productivity, and human-centered effects. It avoids factors that would bring uncertainty in measurement and overcomplicate calculations on ROI. Therefore, remaining elements determine the direct impact of AI within organizations and produce more reliable results, reflecting the overall success of technologies.

### 6.2 Hypothetical Use Case

To illustrate how the validated ROI framework can be applied in practice, consider a hypothetical scenario in which the recruitment department of a medium-sized organization adopts an AI-powered candidate-screening agent. The tool analyzes incoming applications, summarizes profiles, and ranks candidates for hiring managers. Following the logic of the validated framework, the organization assesses productivity gains and financial benefits, while also accounting for human adoption and the fit between its technological maturity and the complexity of the AI solution. A detailed description of how each element translates into measurable variables is presented in this section.

Assume that after one month of using the AI system, the organization identifies two categories of benefits. The first type includes productivity-based benefits in terms of saved time and fewer errors. They can be converted into monetary value using the following formula:

Bprod = (Hours Saved * Fully Loaded Hourly Cost)

For example, the company observes that AI helps to reduce manual screening time and rework, saving around 800 hours per month. Given that employee cost per hour is €50, operational gains generated, Bprod, are 800 * €50 = €40,000/month.

The second type is financial benefits that arise from cost avoidance and opportunity cost from reallocating HR specialists to higher-value tasks such as training and employer branding. To determine their monetary impact, the following calculation can be made:

Bfin = (Cost Avoided + Revenue Uplift)

The revenue generation, Bfin, is €60,000/month as the company avoids hiring costs of €35,000 to third parties since AI reduces reliance on external screening providers. Additionally, it makes €25,000 due to improved funnel management and efficient execution.

Besides benefits, the management also observes its monthly total cost of AI ownership by summing up expenses:

$C_{total} = \Sigma$ (All Costs)

For example, it spends €13k on licensing, €8k on integration, €4k on maintenance, and €10k on process redesign, making a total cost, Ctotal, of €35,000/month.

Next, the company analyzes human adoption and motivation to use the AI system by conducting employee surveys. The statistics demonstrate that HR staff actively engage with the system, trust its recommendations, and feel motivated by being freed from repetitive administrative work. Accordingly, the score for adoption rate is 0.8, while motivation, trust, and usage frequency are 0.7 each. These results can be added to the following formula to measure the overall score:

$A_{human} = \Sigma$ (Human Engagement Indicators) / N

The calculation shows high adoption and motivation with AI, $A_{human} = (0.8 + 0.7 + 0.7 + 0.7)/4 = 0.725$.

At the same time, the organization evaluates the technological fit between its infrastructure and the AI solution. Because the HR data is already hosted on a cloud platform and prior digitalization efforts ensured sufficient processing capacity, the organization's maturity aligns well with the moderate computational demands of the agent. By measuring these factors on a scale of 0 to 1,

organizational maturity is set at 0.7 and AI agent complexity is 0.8. The overall index can be calculated using:

Mfit = 1 - |Maturity Score - Complexity Score|

Therefore, the maturity-complexity fit, Mfit, is estimated at 1 - $|0.7 - 0.8|$ = 0.9. All key variables are then input into the validated ROI formula,

$$\text{ROI}_{\text{AI}} = \frac{(40{,}000 + 60{,}000) \times 0.7 \times 0.9 - 35{,}000}{35{,}000} = 0.8$$

The resulting ROI of 0.8 shows that, under these conditions, the AI agent yields an 80% return relative to its monthly total cost of ownership. To demonstrate the importance of human and contextual factors, the organization also considers a less favorable scenario in which adoption and maturity are lower. For example, $A_{human} = 0.5$ and $M_{fit} = 0.8$, while all other values remain unchanged:

$$\text{ROI}_{\text{AI}} = \frac{100{,}000 \times 0.5 \times 0.8 - 35{,}000}{35{,}000} = \frac{40{,}000 - 35{,}000}{35{,}000} \approx 0.143$$

In this case, the ROI becomes 0.143, indicating a 14.3% return despite identical productivity and financial benefits. This contrast illustrates the core insight of the validated framework: meaningful ROI from AI tools in managerial and decision-making contexts depends not only on direct efficiency and financial gains, but also on whether employees adopt the system and whether the organization's technological maturity is well matched to the complexity of the AI solution.

This approach can be applied to a second hypothetical case where a small-sized company decides to implement a customer service agent. The total cost of integration and maintenance,

Ctotal, amounts to €4,000 per month. The management observes that the AI system speeds up communication with customers, leading to less errors and higher satisfaction. This allows the company to save costs on hiring new employees and increase its revenue, Bfin, by €12,000/month. Along with financial improvements, it notices that productivity becomes higher as the AI agent can address multiple requests in a short time. By converting time savings into value, the company increases operational efficiency, Bprod, by €5,000/month. Then, it analyzes employee level of engagement and motivation to use technology in the workplace. It reviews AI-related requests in the helpdesk and conducts supervisor assessments on the frequency of employee interaction with the agent. It finds a moderate adoption rate of 0.6 and usage frequency of 0.5, also its trust and motivation are both 0.4. The overall score of these values, Ahuman, results in 0.475. Although employee willingness is low, the organization's maturity aligns with the complexity of the agent, as it uses a simple system while being less mature. Thus, the AI complexity is scaled at 0.4, and organizational maturity is 0.3, making a total index, Mfit, of 0.9. To calculate ROI, these values can be put into the final formula:

$$ROI_{AI} = \frac{(12{,}000 + 5{,}000) \times 0.4 \times 0.9 - 4{,}000}{4{,}000} = 0.5$$

The result of 50% demonstrates that the use of an AI agent in a small-sized company is advantageous, but its investment return potential is not fully realized. If more employees were motivated and willing to adopt AI frequently, then the ROI would increase. For example, if Ahuman were 0.5 with other values remaining unchanged, returns would yield 0.9 or 90%. This underscores once more the importance of implementing technological solutions that are both useful for employees and suitable for the development of the organization.

**Individual Contribution:**

## 3 Literature Review

The literature review adopts a semi-systematic approach, combining structured database searches with an integrative synthesis of results (Snyder 2019). This approach makes it possible to identify and compare existing methods for assessing AI productivity and ROI, evaluate their suitability for management-level AI agents, and point out research gaps that can guide the development of a new evaluation framework.

To collect the literature, a semi-systematic search was conducted across Google Scholar, SSRN, MDPI, ScienceDirect, and Springer. The search used keywords *AI, AI Tools, ROI, Process Optimization, Workflow Automation,* and *Human-AI Collaboration*. Only English-language studies published between 2022 and 2025 were considered, focusing mainly on peer-reviewed and data-driven research, including field and controlled experiments as well as papers outlining future research agendas. In total, twenty articles were selected for systematic review and thematic analysis.

### 3.1 AI Productivity Gains

Research shows that AI can boost both individual and organizational productivity, especially in tasks that involve information processing or decision support (Noy and Zhang 2023; Cui et al. 2025; Brynjolfsson, Li, and Raymond 2025). In management settings, AI agents are increasingly used to assist with planning, resource allocation, and coordination, areas that have traditionally depended on human judgment (Islam et al. 2024; Sauer and Burggräf 2024; Yang, Lovett, Li, and Hou 2025).

The productivity of organizations increases with AI adoption, enabling employees to efficiently execute tasks. The evidence comes from experiments that are set up to test the capabilities of GenAI assistants. For example, research in the customer service setting shows that AI can analyze a large volume of data, helping to support over 5,000 customers (Brynjolfsson, Li, and Raymond 2025, 891). This suggests that AI technologies help to resolve a greater number of issues, leading to positive feedback. As a result, the company benefits from improved communication and employees gain better experience.

Another advantage of AI is its broad capabilities, meaning that it can be applied for multiple purposes in various types of industries. For example, digital tools can assist with technical or cognitive tasks. In the case of programming, technologies such as GitHub Copilot support workers in writing efficient codes (Peng et al. 2023, 5). This guarantees acceleration of the coding process as programmers spend less time on its development. A similar outcome in terms of time and quality is observed in the case of customer support. Employees using ChatGPT tend to improve clarity of their responses and their planning process in professional writing (Noy and Zhang 2023, 3). The two cases highlight common productivity gains which include time, quality, and worker experience (Al Naqbi, Bahroun, and Ahmed 2024, 5). Moreover, they demonstrate the effectiveness of AI in occupations with repetitive tasks.

In addition to analysis of the efficiency outcomes, some studies propose frameworks for ROI computations. One of the examples includes a calculator specifically designed to measure the value of AI in radiology (Bharadwaj et al. 2024, 1683). The experiment, conducted over five years, shows that AI technology supports medical staff by reducing time spent on diagnostic tasks. By integrating these data into the ROI calculator, results suggest that AI optimizes processes, leading to higher financial returns.

Another experiment uses a cost model to evaluate ROI of an AI tool in customer service (Katragadda 2024, 136). Specifically, it aims to understand how financial outcomes vary with operational changes from introducing a chatbot. For instance, whether increased returns are associated with improved resolution rates. The findings reveal that AI enhances overall productivity, and the cost model correspondingly indicates an increase in ROI. This means that the chatbot reduces operational costs, thereby contributing to the financial performance of an organization. Nevertheless, this is an industry-specific case with measurements of certain metrics. Therefore, the cost model provides limited representation of its usage in broader managerial contexts.

Overall, the implementation of AI generates several benefits such as research and development, innovation, and productivity growth (Filippucci et al. 2024, 16-18). However, it is important to identify key factors that drive the firm's success from AI investments. For instance, organizations should use technologies to create new products or improve their existing processes (Babina et al. 2024, 15-19). This, in turn, will add more value and ultimately result in revenue generation. To quantify efficiency gains, it is necessary to develop a comprehensive ROI framework by investigating different existing approaches.

### 3.2 ROI Methodologies

While there are several ways to measure the ROI of AI, few studies focus on management-level agentic AI, where success is not only about efficiency but also about better coordination, decision quality, and alignment with organizational goals. According to recent empirical research, methodological approaches can be divided into four categories.

**Experimental and quasi-experimental** methods are widely used by researchers to study productivity effects of AI tools. Experimental designs usually adopt randomized controlled trials (RCTs), meaning that a random group of people gain access to technology (Brynjolfsson, Li, and Raymond 2025, 901-902). The purpose is to observe its effect on employee performance, after which the tool is gradually introduced to other groups, known as staggered rollouts. The final step of the procedure is to compare results and quantify key performance indicators (KPIs) such as task completion rates (Peng et al. 2023, 4). In contrast, quasi-experimental designs use existing data to get a full overview of how technology is being used inside organizations (Al Naqbi, Bahroun, and Ahmed 2024, 3). This allows researchers to summarize and compare the impact of AI across multiple industries. Thus, both methods can be useful for short-term analysis, but they are incapable of performing long-term assessments due to their limited ability to capture evolving managerial and value-creation effects.

**KPI-based** approaches are used to capture real-time impacts of AI on the performance of organizations. The analysis includes a set of metrics such as handle times, customer satisfaction, or brand reputation (Katragadda 2024, 135-136). It is necessary to collect data and monitor outcomes, for example, through customer feedback or reports. This allows companies to obtain both qualitative and quantitative results, which can be further integrated into the ROI formula. Nevertheless, such methods have several limitations, including data reliability, time lag, and scalability constraints, according to Katragadda (2024).

**Cost-benefit** analysis offers a more detailed assessment of AI impacts through comparison of its costs and benefits. This can effectively be measured using a ROI calculator, as demonstrated in the study with AI in radiology (Bharadwaj et al. 2024, 1680-1683). The cost components typically include installation and service fees, while recorded benefits are time savings and scaling

opportunities. The outputs indicate changes in revenue and accuracy rates, as well as in workflow processes. However, the challenge of such an approach is that its data primarily depends on expert beliefs and overlooks indirect impacts.

**Hybrid frameworks** advance ROI formulas by considering additional elements which may influence returns. Besides technical metrics, they account for human-centered, contextual, and temporal dimensions (Meimandi et al. 2025, 6). This suggests that formulas track user feedback, measure stability, and evaluate potential risks of AI systems. Therefore, they provide a better reflection of real-time impacts on organizational performance. To capture these changes in more detail, organizations may choose to monitor each step of AI innovation lifecycle (Ganguly and Patra 2025, 10-12). They can also deploy multiple AI agents to optimize processes and maximize ROI. However, this is a complex approach, requiring ongoing oversight and adaptability to external factors. Studies using hybrid frameworks usually demonstrate theoretical applications and predict likely outcomes.

Taken together, these four approaches reveal a clear trade-off between causal precision and financial relevance. Experimental designs identify short-term efficiency effects with strong validity, while cost–benefit and hybrid models estimate ROI but depend on assumptions or modeled data. Integrating both perspectives could provide a more complete understanding of AI's overall organizational impact.

### 3.3 Practical Frameworks

As illustrated in several case studies, these methodologies can be effectively applied in practice. For example, researchers carry out an experiment using Randomized Controlled Trials (RCT) to measure the impact of GenAI in the software development sector (Cui et al. 2025, 16-

20). They select a group of developers at Microsoft and Accenture, to whom they grant access to a coding assistant. During the course of the study, they report improvements in employee productivity. Specifically, they find that AI supports less-experienced workers in completing tasks effectively. However, another research presents an opposing effect, claiming that technology has limited benefits on lower-skilled employees (Otis et al. 2024, 31-33). The reason is that they lack sufficient technical knowledge to use AI, especially in case of complex tasks. This highlights the need to adopt AI systems according to different skills of workers to ensure their effectiveness.

Another study describes a KPI-based approach to determine the impact of AI in fraud detection and risk management (Islam et al. 2024, 5-6). It analyzes quantitative metrics based on statistical data from financial reports and company datasets. For example, it tracks detection rates before and after AI implementation, as well as its economic impact. In addition, qualitative data is gathered through expert interviews, providing insights into ethical implications and internal controls. The analysis of KPIs reveals reduced errors and improved detection accuracy, resulting in overall cost savings. This suggests that organizations are able to expand their operations due to the generation of higher revenues.

An alternative study uses cost-benefit analysis to test the impact of an AI system in the manufacturing sector (Alam, Islam, and Shil 2023, 556-559). It collects the necessary data and converts it into monetary value to understand financial outcomes. Reported benefits include downtime reductions, improved equipment uptime, and productivity gains. Costs are categorized into implementation, training, and operational expenses. When applied to a ROI formula, the analysis reveals favorable returns, suggesting that AI optimizes processes and saves costs of the organization.

Lastly, a hybrid framework is introduced in the study exploring the role of AI in the case of production management (Sauer and Burggräf 2024, 529-531). It intends to find a balanced collaboration between humans and AI systems to optimize workflow. For example, it measures several variables such as data availability, error rates, and decision complexity. The procedure is divided into four phases, where tasks are analyzed at each step. This method allows organizations to create comprehensive frameworks that complement human skills. As a result, they can facilitate their decision-making process and achieve operational gains.

Additional studies demonstrate that adopting multiple AI agents can improve company performance and contribute to framework development. These technologies support in tracking individual employee capabilities and allocate tasks according to skill levels (Yang, Lovett, Li, and Hou 2025, 18-20). This enables organizations to establish an effective plan for their workforce and adapt to real-time changes in behavior. Therefore, they can observe measurable impacts on ROI and achieve overall productivity gains.

### 3.4 Impact Assessment Framework

To assess the impact of various AI solutions on organizational performance, the framework proposed by Bornet et al. (2025, 263-265) can be applied. On a five-point scale, it measures four key criteria: time saved, strategic value of freed time, error reduction potential, and scalability impact. A higher total score indicates a greater contribution of AI along this scale:

<10: Limited impact

10-13: Moderate productivity gains

14-17: Strong operational improvement

18-20: Transformational impact

Table 1 offers a detailed description of each criterion and its score, including a list of metrics to ensure the objective assessment. Furthermore, it outlines cases in which measuring each criterion is useful and identifies suitable evaluators for conducting these assessments.

| **Criteria** | **Description** | **Scores (1-5)** | **Metrics for Objective Assessment** | **Use Cases** | **Evaluators** |
|---|---|---|---|---|---|
| Time Saved | Measures how much time AI saves per task | 1: Less than one hour per week<br>2: 1-3 hours per week<br>3: Several hours per week<br>4: 1-2 hours per day<br>5: Multiple hours daily | Task time reduction, task completion rate, degree of multitasking | Repetitive tasks | Process engineers |
| Strategic Value of Freed Time | Evaluates how freed time contributes to strategic work | 1: Limited value (routine tasks)<br>2: Some analytical tasks<br>3: Moderate value (mix of routine and analytical tasks)<br>4: Operational tasks<br>5: High value (strategic tasks) | Completion rate of high-value tasks, opportunity cost of manual work, impact on strategic KPIs | Management and customer service contexts | Business analysts |
| Error Reduction Potential | Measures how effectively AI reduces errors | 1: Minimal errors<br>2: Occasional errors<br>3: Frequent errors moderately reduced<br>4: Significant errors<br>5: Frequent errors eliminated | Error rate before/after AI use, accuracy improvement, type of errors prevented (human or systematic) | High-precision industries (medical, defense) | Quality assurance engineers |
| Scalability Impact | Evaluates how likely AI can be used across organization | 1: Single-use task<br>2: Departmental<br>3: Moderately repeatable across teams<br>4: Cross-functional<br>5: Organizational | Process automation, number of users, replication cost and time | Pilot projects | Strategic planners, IT managers |

**Table 1.** AI Impact Assessment Framework (adapted by Bornet et al. 2025)

For illustration, the framework can be applied to the study by Bharadwaj et al. (2024), which investigates the impact of an AI-powered radiology platform. Their findings show that the technology reduces patient waiting and scanning times, saving a total of 145 days over a five-year period (Bharadwaj et al. 2024, 1682). The available time allows radiologists to carry out additional scans and procedures, leading to a higher number of diagnoses. This facilitates a restructuring of work schedules and emphasizes complex cases, since the AI tool assists in reprioritizing medical issues. Moreover, it improves diagnostic accuracy through better detection of positive cases. For example, AI-supported workflows increase the identification of patients with incidental pulmonary nodules by 21%, which is clinically significant as such nodules can indicate early-stage lung cancer (Bharadwaj et al., 2024, 1680). The tool becomes a valuable solution for the organization, benefiting doctors, administrators, and IT teams. Furthermore, the system requires minimal retraining and maintenance, reducing implementation time and associated costs. To sum up, the AI-powered radiology platform saves multiple hours daily (5), enhances strategic clinical work (5), reduces significant diagnostic errors (4), and scales organization-wide (5). The total score adds up to 19, making a transformational impact as it improves both productivity and quality, ultimately contributing to a substantial increase in ROI.

Another example of applying the framework can be demonstrated using the study by Brynjolfsson et al. (2025), which tests the effects of GenAI conversational assistants in customer support. The system reduces the average handling time by eight minutes per chat, saving roughly one hour per day (Brynjolfsson et al. 2025, 903). This enables agents to manage a higher volume of interactions and engage in multiple conversations simultaneously. Beyond completing routine tasks, the AI tool examines product information as well as employee and customer behavior to offer the most suitable solutions. Although this contributes to higher customer satisfaction, human

oversight remains necessary to perform complex tasks and intuitive situations. Moreover, performance outcomes vary depending on worker experience levels, specifically benefiting less-experienced employees by improving their grammar mistakes and English fluency. This limits the standardization of the AI system across teams and requires adjustments in order to benefit all employees. Overall, the GenAI conversational assistant saves an hour daily (4), performs both routine and analytical tasks (3), but reduces only occasional errors (2), and scales across a single department (2). The resulting total score of 11 implies moderate productivity gains, as the technology improves efficiency but does not change strategic processes.

The two cases highlight that AI generates the highest impact on data-intensive tasks, whereas its impact remains limited in the case of novel tasks. The systems fail to effectively solve unfamiliar issues due to insufficient creativity, resulting in the need for human-AI collaboration to drive organizational change. Additionally, it is important to consider the specific challenges of using AI agents in managerial roles, where success depends on how well these systems support human judgment, help organizations learn, and improve coordination.

### 3.5 Current Limitations in the Literature

Although existing literature establishes a relationship between AI adoption and ROI, it also emphasizes several limitations. Most studies adapt their methodologies to specific industries rather than proposing a generalized solution. They demonstrate a limited impact of AI, for example using a digital assistant exclusively for developer tasks (Peng et al. 2023, 7-8). Thereby, overlooking its potential contribution to other organizational roles such as project management. Furthermore, studies usually focus on short-term analysis, which fails to capture indirect effects (Noy and Zhang 2023, 12). To provide long-term assessment, future research and additional data are required.

Additionally, there are uneven effects of AI inside organizations, and employers should be aware of possible risks on output quality and employee knowledge. Although AI is capable of performing human work, it may cause a decline in overall performance (Dell'Acqua et al. 2023, 17-19). For example, worker experience can decrease due to delegation of tasks to AI systems. This means that employees are deprived of gaining expertise and developing new skills. Moreover, interns receive fewer assignments which expose them to incomplete on-the-job training. Another risk involves digital trust as employees may blindly accept responses provided by AI. The failure to validate information results in a greater probability of errors in the workplace. This further suggests that employees disregard important data and deteriorate their cognitive abilities (Dell'Acqua 2022, 28-30). Hence, instead of replacing human labor, AI should complement their skills to enable organizational growth.

A further limitation is that studies rarely assess the financial impact of AI systems, rather they focus on operational outcomes. For instance, they provide evidence that coding assistants enhance efficiency, but do not express these benefits in financial terms (Cui et al. 2025, 19-20). Although there are studies that apply ROI frameworks, they usually represent limited and industry-specific cases. This emphasizes the need to develop a general methodology that combines both outcomes. It would offer a more detailed understanding of the various impacts of AI tools. To ensure its results are complete and objective, the data should constantly be verified (Islam et al. 2024, 17).

Finally, it is important to note that AI systems are highly complex with different degrees of organizational readiness (Heimberger, Horvat, and Schultmann 2024, 13-14). Therefore, advanced frameworks are particularly useful in realizing the full potential of ROI. Beyond technical metrics, they usually incorporate economic, safety, and human-centered dimensions

(Meimandi et al. 2025, 6). This increases the precision of hybrid models, as they become adaptable to external changes while pursuing organizational objectives (Okunola, Ahsun, and Blace 2025, 9-11). Theoretically, they can also be scaled across the organization, but further practical validation is required. These gaps highlight the need to design a comprehensive approach to determining the ROI of AI agents within organizations.

Overall, the literature shows that AI can improve productivity, but there is still no consistent way to assess both its operational and managerial impact. Most existing approaches focus either on technical efficiency or on financial returns, while paying less attention to how agentic AI changes managerial work, decision quality, or employee capabilities. This gap highlights the need for a broader ROI framework that captures the realities of management-level agentic AI and human-AI collaboration. Table 2 offers an overview of various methods used by researchers to perform measurements of AI productivity and ROI, summarizing key findings and limitations in each specific case.

| Author & Year | Sector / Context | Method Type | Main Metrics | Key Findings | ROI Link / Limitation |
|---|---|---|---|---|---|
| **Brynjolfsson et al. 2025** | Customer Support | Experimental (RCT) | Issues resolved / hour, satisfaction | +14–15 % productivity gain, higher retention | Indirect ROI via productivity; short-term focus only |
| **Noy and Zhang, 2023** | Professional Writing | Experimental | Task time, quality of output | Higher speed and quality for cognitive tasks | No financial conversion; individual-level only |
| **Peng et al. 2023** | Software Development | Quasi-Experimental (GitHub Copilot) | Coding time, completion rate | Time savings ≈ 55 %, improved developer efficiency | ROI not quantified; sector-specific |
| **Bharadwaj et al. 2024** | Healthcare (Radiology) | Cost–Benefit / ROI Model | Workflow time, installation costs vs benefits | Workflow improvement → financial returns over 5 years | Strong ROI focus but context-bound |
| **Katragadda, 2024** | Customer Service | KPI-Based Model | Resolution rates, handling time | Operational metrics converted to ROI model | Scalability and data-quality issues |
| **Alam et al. 2023** | Manufacturing | Cost–Benefit Analysis | Downtime, training costs | Predictive maintenance ↑ productivity, ↓ costs | ROI calculated; limited to process optimization |
| **Sauer and Burggräf, 2024** | Production Management | Hybrid Framework | Human–AI collaboration level, error rates | Defines four-phase collaboration model | Conceptual validation needed |

| | | | | | |
|---|---|---|---|---|---|
| **Ganguly and Patra, 2025** | Media Decisioning | Hybrid / Multi-Agent Framework | Lifecycle ROI indicators | Integrates AI agents across business processes | Mostly theoretical; no empirical test |
| **Okunola et al. 2025** | Workforce Transformation | Managerial ROI Framework | Skill metrics, efficiency indicators | Links AI adoption to human ROI factors | Exploratory; needs validation |
| **Filippucci et al. 2024** | Cross-Industry / OECD | Macro Data Analysis | Productivity, growth indicators | AI positively linked to growth & innovation | Macro level; no firm ROI mechanism |

**Table 2.** Overview of methods and approaches used to measure AI productivity and ROI.

# 5 Main Findings and Recommendations

## 5.1 Sampling Rationale

To validate the theoretically derived framework, this study uses a purposeful sampling technique (Bouncken et al. 2025, 4). It selects experts in the field who have relevant knowledge and experience in the research topic. They provide an in-depth explanation of the framework elements and their applicability to real organizations. Therefore, valuable insights and recommendations are gathered even with a small number of individuals. The numerical threshold is of less importance in qualitative research since its primary focus is on data richness (Ahmed 2025, 3). It ensures that enough information is available to present various themes and enable discussion of its patterns and differences (Moser and Korstjens 2018, 11). Consequently, four expert interviews were conducted to evaluate the framework and understand adjustments to be made.

## 5.2 Participant Descriptions

**Participant 1** – Professor, co-owner and managing partner

The first participant is a professor at university as well as a co-owner and managing partner of a consulting firm. The company focuses on talent management, but it started off with software development and used algorithm-based machine learning. He has previously worked with AI on a large language model and predictive and analytical tools.

**Participant 2** – Professor, chairman and non-executive director

The second interviewee is also a university professor, a chairman of two companies and a non-executive director of a third one. He used to be a vice CEO of a bank responsible for internal

processes and had executive roles in the area of planning and control for 12 years. He operated with neural networks to power analytical tools and robots to manage workflows in the banking industry.

**Participant 3** – CEO

The third participant is a CEO of a consulting firm, engaged in management, sales, and high-level discussions with clients on AI since the company is dedicated to digital transformation and enterprise project management. It initially used OpenAI on enterprise scenarios which was then replaced by agents to automate processes.

**Participant 4** – Professor

The fourth interviewee has a background in electrical engineering and computer science. He currently teaches students about product design and development, as well as innovation. In the past, he was a board member of a media company discussing the usefulness of AI in the workplace.

Overall, participants have experience in holding high-level management positions and delving into the topic of AI. They have solid backgrounds in organizational development and leadership, as well as consulting and academic expertise. Their theoretical and practical knowledge of automation technology contributes to enrichment of the findings. This provides valuable feedback about the theoretically derived framework which is used to create a validated post-interview framework.

### 5.3 Interview Feedback

Starting with the questions about the **model flow**, participants commented that they find it comprehensive on a generic basis. Upon closer inspection, there could be some adjustments or

variables required, specifically for complex use cases. One of the participants mentioned that there is no need to specify the type of processes since it would already be interpreted in the formula. Nonetheless, the rest agreed that each layer is equally relevant and should be kept in the model.

During the discussion of the **indicator list**, most of the interviewees agreed that all components are fundamental. Every indicator depends on each other, specifically highlighting the importance of integration and deployment in the input section. However, in the case of AI, data preparation and IT infrastructure might be unnecessary. Typically, companies will operate on cloud systems except if they belong to the medical industry. This suggests that the choice of components changes according to different business sectors.

Participants further highlighted the difficulties that may arise when measuring the list of elements. For instance, the accuracy of outputs depends on the repetition of the process. If the company manufactures a product frequently, it can easily determine the amount of time and material required. Whereas, with a one-off product, it is difficult to maintain standardization and detailed information to guarantee an efficient result. Therefore, measuring outcomes becomes complicated and requires a lot of assumptions. This makes it harder to complete a final evaluation of ROI.

In terms of **potential additional indicators**, participants suggest including opportunity cost and throughput efficiency. This means assessing the changes in the quality of work that AI tools can bring compared to humans. For example, technologies can analyze tons of information and react faster than people. Another element to consider is time taken to redesign the processes of organizations deciding to implement AI. This includes the cost of resources required today to produce the process. This also requires hiring specialists to explain the impact of AI adoption to

company representatives. The final recommendation is to divide the types of processes into generative or classic, depending on the need for reasoning and interpretation or if it is simply a workflow with calculations.

The questions on the **ROI framework** show that interviewees consider the hybrid approach as comprehensive. They point out that the adoption rate is particularly relevant since lots of AI impact narrows down to a human factor. One of the responses (see Appendix B, Participant 2) states that "if people are not open to using AI in their processes, that will have a huge impact on the ROI. So, it depends on how people work with or work against AI." The adoption rate also ensures that the tool is useful to people, helping them save time and generate money. While some say that this element is easy to measure, others argue that it depends on the department and application. Its feasibility can vary whether it is applied for an entire organization or process-by-process. Regarding other framework elements, productivity benefits are easy to measure, and financial benefits need some assumptions. The overall estimation depends on the industry and the practice of management. If the company has complete information, it becomes easier to integrate automation and assess productivity gains. Whereas aggregate or incomplete data may hinder measurements of AI impact.

For potential **missing parts of the framework**, participants recommend adding the maturity level of the organization in terms of digital development. This reflects the openness and willingness of people to start working with digital tools. Adoption rate should also capture human capability of using AI for non-autonomous models in terms of training and trusting the system. This will have a significant impact on output, yielding different results. In combination with human engagement and motivation, the framework can provide a perception. It can allow companies to understand their employees better and get rid of non-value adding tasks. Furthermore, including

an opportunity for increased quality can demonstrate new possibilities that become available with AI adoption. A final recommendation is adding a temporal variable as its relevance might change in the future. It relates to brand reputation, showing that AI tools can bring benefits or damages to organizations due to their uncertainty and novelty.

Regarding the practical **challenges of the ROI framework**, participants emphasize that it would be difficult to estimate values before investing in AI. The decision would be based on assumptions, whereas measuring the consequences of AI investment would be easier. However, these effects would vary according to the complexity of agents. For example, the use of simple systems typically brings less benefits and requires specialized services such as APIs (Application Programming Interfaces). Multi-agent systems tend to improve work efficiency and even perform human tasks. Although this is beneficial for companies, it might bring difficulties in assessing the cost of people. If the company decides to replace some of its employees with AI, it must pay severance fees. Depending on the country, this could be a reasonable approach since the severance package would be lower than the total cost of salaries. When it comes to performing ROI calculations in such a case, it becomes hard to measure the impact of people and productivity cycles. To identify overall savings and analyze multiple variables, consultant expertise could be required. It could involve the use of other financial metrics such as IRR (Internal Rate of Return) to carry out complex computations that incorporate time variables.

Furthermore, companies should be aware of potential **risks of AI tools** and control the operational process of systems. They should detect technical errors and fix them in a timely manner. This will help to minimize the probability of reputational damage caused by technology. They should understand that AI tools are unable to tackle all sorts of problems. Thus, they should carefully analyze their decisions to take on new risks, avoiding overconfidence. Ultimately, the

adoption of AI within the organization is a choice made by humans who therefore also have to face potential consequences.

Proceeding with the **overall assessment** of the ROI framework, different opinions were expressed on its feasibility score. On a scale of 1 to 10, the first participant gave a rating of 8 due to difficulty in measuring outcomes. He suggested accounting for human engagement with AI to further estimate ROI from an emotional and motivational perspective, rather than financial. The second participant rated the framework as 9 for sophisticated companies that have detailed and timely information. The complete data will allow managers to better understand the usefulness of AI tools and fully benefit from them. The third interviewee said that the score depends on the type of implementation. Particularly, he gave a 6/7 if AI impact is measured for a single process, or a 5 for entire organization. Although the process-by-process method is easier to calculate, it requires specialists to implement AI and connect it to systems. The fourth interviewee commented that the framework is totally applicable, giving it a rating of 10. However, he assumes that companies usually decide not to thoroughly follow guidelines, lowering score to 5.

Final questions on the impact assessment table provide insights into how experts would evaluate the **impact of AI tools**. They have agreed that measuring time, strategic value, errors, and scalability is important, especially from a technical perspective. The strategic value of freed time provides opportunity cost, which is an efficient indicator of process utility. Cost efficiency provides detailed analysis of the workforce and input reductions. On a managerial level, it could be interesting to add the element of money to demonstrate the budget impact of AI. In that case, error reduction would be divided into two impacts, time and money. In addition, a minor adjustment to be made is to replace the term strategy with value-added activities. Strategy is

commonly overused and misinterpreted, thus it is better to use strategic to describe different activities, and tactical to describe the same activities but with better value.

### 5.4 Themes and Recommendations

Having gathered feedback from all interviews, it can be grouped into key themes such as model flow, indicator list, ROI framework, challenges, as well as overall and impact assessments. Table 3 offers an overview of relevant recommendations given by experts in relation to each topic. The types of feedback outline statements that either confirm the existing approach or emphasize critical information and enhancement-oriented changes. This will be used to design the post-interview ROI hybrid framework to increase its feasibility and reliability.

| Themes | Feedback | Types |
|---|---|---|
| **Model Flow** | Generally comprehensive | Confirmatory |
| **Indicator List** | All fundamental | Confirmatory |
| | Difficult to estimate accuracy of outputs | Critical |
| | Data preparation and IT infrastructure sometimes unnecessary | Critical |
| | Add opportunity cost and throughput efficiency | Enhancement |
| | Add time spent on process redesign, including costs of resources and specialists | Enhancement |
| | Divide types of processes into classic or generative | Enhancement |
| **ROI Framework** | Comprehensive hybrid approach, adoption rate is relevant | Confirmatory |
| | Add maturity level of the organization in terms of digital development | Enhancement |
| | Within adoption, measure human capability of using AI for non-autonomous models | Enhancement |
| | Add a measure of human engagement and motivation | Enhancement |
| | Add opportunity for increased quality | Enhancement |
| | Add brand reputation | Enhancement |
| **Challenges** | Difficult to estimate values before investing in AI, assumptions required<br>Difficult to measure outcomes and therefore ROI<br>Benefits vary with the complexity of agents | All critical |

| | | |
|---|---|---|
| | Difficult to assess the cost of people and overall savings | |
| | Requires consultants and complex computations | |
| | Risk of reputational damage caused by AI | |
| | Overconfidence on AI tools | |
| | Account for human engagement to grasp emotional and motivational perspective | Critical |
| **Overall** | Feasible if companies have detailed and timely information about their processes | Confirmatory |
| **Assessment** | Easier to measure for process-by-process implementation than entire organization | Critical |
| | Applicable unless companies choose not to follow guidelines | Confirmatory |
| | Comprehensive assessment | Confirmatory |
| **Impact** | Add element of money to measure the budget impact of AI | Enhancement |
| **Assessment** | Evaluate error reduction in terms of money and time | Enhancement |
| | Rename strategy to value-added activities | Critical |

**Table 3.** Expert recommendations about ROI framework and its key themes

# 8 Conclusion

This thesis presented the development of the framework for organizations to assess ROI of their AI systems in managerial and decision-making contexts. It addressed the key gap found in literature, which was the absence of an evaluation tool that would simultaneously measure the financial and operational effects of AI. Such a hybrid approach became the focus of the current work and formed the basis for the pre-interview version of the ROI formula. It aimed at identifying the fundamental components that provide realistic measurements and reflect the real-time impact of AI. To verify credibility and coherence of the existing method, this study performed a qualitative analysis in the form of expert interviews. The discussion highlighted the importance of human engagement with AI and complexity of systems that match the maturity level of companies. These elements, along with quantitative metrics, were incorporated into the post-interview formula, capturing four dimensions: financial, productivity, human-centered, and contextual. This allowed the hybrid framework to adapt to individual characteristics of companies and interpret their relationship with technologies. Thereby offering a dynamic assessment of AI impacts and demonstrating significant changes in returns. Consequently, the refined approach was characterized as practical and theoretically founded but requires further validation through real-world cases. Organizations should test the applicability of the framework within their processes over a longer period of time. This would strengthen the analysis and emphasize sectors where the framework is specifically useful and measurable. Nevertheless, the proposed tool provides managers with an understanding of the financial and economic impact of AI systems, contributing to their investment decisions and future success of organizations.

**Reference List**

Alam, Mahfuz, Md Rafiqul Islam, and Sanjib Kumar Shil. "AI-Based predictive maintenance for US manufacturing: reducing downtime and increasing productivity." *International Journal of Advanced Engineering Technologies and Innovations* 1, no. 01 (2023): 541-567, https://ijaeti.com/index.php/Journal/article/view/711

Al Naqbi, Humaid, Zied Bahroun, and Vian Ahmed. "Enhancing Work Productivity through Generative Artificial Intelligence: A Comprehensive Literature Review" *Sustainability* 16, no. 3 (2024): 1166, https://doi.org/10.3390/su16031166

Ahmed, Sirwan Khalid. "Sample size for saturation in qualitative research: Debates, definitions, and strategies." *Journal of Medicine, Surgery, and Public Health,* no. 5 (2025): 1-6, https://doi.org/10.1016/j.glmedi.2024.100171

Babina, Tania, Anastassia Fedyk, Alex He, and James Hodson. "Artificial intelligence, firm growth, and product innovation." *Journal of Financial Economics* 151 (2024): 103745, https://doi.org/10.1016/j.jfineco.2023.103745

Bharadwaj, Prateek, Lauren Nicola, Manon Breau-Brunel, Federica Sensini, Neda Tanova-Yotova, Petar Atanasov, Franziska Lobig, and Michael Blankenburg. "Unlocking the value: quantifying the return on investment of hospital artificial intelligence." *Journal of the American College of Radiology* 21, no. 10 (2024): 1677-1685, https://doi.org/10.1016/j.jacr.2024.02.034

Birt, Linda, Suzanne Scott, Debbie Cavers, Christine Campbell, and Fiona Walter. "Member checking: a tool to enhance trustworthiness or merely a nod to validation." *Qualitative health research* 26, no. 13 (2016): 1802-1811.

Brinkmann, Svend. “Qualitative interviewing.” *Oxford university press*, 2013.

Bryman, Alan. “Social research methods.” *Oxford university press*, 2016.

Brynjolfsson, Erik, Danielle Li, and Lindsey Raymond. "Generative AI at work." *The Quarterly Journal of Economics* 140, no. 2 (2025): 889-942, https://doi.org/10.1093/qje/qjae044

Bornet, Pascal, Jochen Wirtz, Thomas H. Davenport, David De Cremer, Brian Evergreen, Phil Fersht, Rakesh Gohel, and Shail Khiyara. “Agentic Artificial Intelligence: Harnessing AI Agents to Reinvent Business, Work, and Life.” *World Scientific Connect* (2025): 572, https://doi.org/10.1142/14380

Bouncken, Ricarda B., Wojciech Czakon, and Florian Schmitt. "Purposeful sampling and saturation in qualitative research methodologies: recommendations and review." *Review of Managerial Science* (2025): 1-37, https://doi.org/10.1007/s11846-025-00881-2

Cui, Zheyuan Kevin, Mert Demirer, Sonia Jaffe, Leon Musolff, Sida Peng, and Tobias Salz. "The Effects of Generative AI on High-Skilled Work: Evidence from Three Field Experiments with Software Developers." *Available at SSRN 4945566* (2025): 38, https://dx.doi.org/10.2139/ssrn.4945566

Dell’Acqua, Fabrizio. "Falling asleep at the wheel: Human/AI Collaboration in a Field Experiment on HR Recruiters." *Laboratory for Innovation Science, Harvard Business School* (2022):

51, https://www.almendron.com/tribuna/wp-content/uploads/2023/09/falling-asleep-at-the-whee.pdf

Dell'Acqua, Fabrizio, Edward McFowland III, Ethan Mollick, Hila Lifshitz-Assaf, Katherine C. Kellogg, Saran Rajendran, Lisa Krayer, François Candelon, and Karim R. Lakhani. "Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality." *Harvard Business School Technology & Operations Mgt. Unit Working Paper*, no. 24-013 (2023): 58, http://dx.doi.org/10.2139/ssrn.4573321

European Commission. "Proposal for a regulation of the European Parliament and the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts." *EUR-Lex-52021PC0206* (2021).

Filippucci, Francesco, Peter Gal, Cecilia Jona Lasinio, Alvaro Leandro, and Giuseppe Nicoletti. "The impact of Artificial Intelligence on productivity, distribution and growth." *OECD Artificial Intelligence Papers*, no. 15 (2024): 63, https://doi.org/10.1787/8d900037-en

Funnell, Sue C., and Patricia J. Rogers. *Purposeful program theory: Effective use of theories of change and logic models*. John Wiley & Sons, 2011.

Ganguly, Kaushik, and Amit Patra. "Designing Multi-Agent Generative AI Framework for ROI-Driven Media Decisioning." *Available at SSRN 5258226* (2025): 20, https://dx.doi.org/10.2139/ssrn.5258226

Gregor, Shirley, and Alan R. Hevner. "Positioning and presenting design science research for maximum impact." *MIS quarterly* 37, no. 2 (2013): 337-355.

Haynes, Stephen N., David Richard, and Edward S. Kubany. “Content validity in psychological assessment: A functional approach to concepts and methods.” *Psychological assessment* 7, no. 3 (1995): 238-247.

Heimberger, Heidi, Djerdj Horvat, and Frank Schultmann. "Exploring the factors driving AI adoption in production: a systematic literature review and future research agenda." *Information Technology and Management* (2024): 1-17, https://doi.org/10.1007/s10799-024-00436-z

Hevner, Alan R., Salvatore T. March, Jinsoo Park, and Sudha Ram. "Design science in information systems research." *MIS quarterly* 28, no. 1 (2004): 75-105.

Islam, Tariqul, S. M. Islam, Ankur Sarkar, A. Obaidur, Rahman Khan, Rakesh Paul, and Md Shadikul Bari. "Artificial Intelligence in Fraud Detection and Financial Risk Mitigation: Future Directions and Business Applications." *International Journal for Multidisciplinary Research* 6 (2024): 1-23, https://doi.org/10.36948/ijfmr.2024.v06i05.28496

Kallio, Hanna, Anna-Maija Pietilä, Martin Johnson, and Mari Kangasniemi. "Systematic methodological review: developing a framework for a qualitative semi-structured interview guide." *Journal of advanced nursing* 72, no. 12 (2016): 2954-2965.

Katragadda, Vamsi. "Measuring ROI of AI implementations in customer support: A data-driven approach." *Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023* 5, no. 1 (2024): 133-140, https://doi.org/10.60087/jaigs.v5i1.182

Legner, Christine, Torsten Eymann, Thomas Hess, Christian Matt, Tilo Böhmann, Paul Drews, Alexander Mädche, Nils Urbach, and Frederik Ahlemann. "Digitalization: opportunity and

challenge for the business and information systems engineering community." Business & information systems engineering 59, no. 4 (2017): 301-308.

Lincoln, Yvonna S., and Egon Guba. “Naturalistic Inquiry.” *SAGE Publications*, 1985.

MacKenzie, Scott B., Philip M. Podsakoff, and Nathan P. Podsakoff. "Construct measurement and validation procedures in MIS and behavioral research: Integrating new and existing techniques." *MIS quarterly* 35, no. 2 (2011): 293-334.

Meimandi, Kiana Jafari, Gabriela Aránguiz-Dias, Grace Ra Kim, Lana Saadeddin, and Mykel J. Kochenderfer. "The Measurement Imbalance in Agentic AI Evaluation Undermines Industry Productivity Claims." *arXiv preprint arXiv:2506.02064* (2025): 15, https://doi.org/10.48550/arXiv.2506.02064

Miles, Matthew B., A. Michael Huberman, and Johnny Saldana. "Qualitative data analysis: A methods sourcebook." *SAGE Publications*, 2014.

Moser, Albine, and Irene Korstjens. "Series: Practical guidance to qualitative research. Part 3: Sampling, data collection and analysis." *European journal of general practice* 24, no. 1 (2018): 9-18, https://doi.org/10.1080/13814788.2017.1375091

Nowell, Lorelli S., Jill M. Norris, Deborah E. White, and Nancy J. Moules. "Thematic analysis: Striving to meet the trustworthiness criteria." *International Journal of Qualitative Methods* 16, no. 1 (2017): 1609406917733847.

Noy, Shakked, and Whitney Zhang. "Experimental evidence on the productivity effects of generative artificial intelligence." *Science* 381, no. 6654 (2023): 187-192, https://doi.org/10.1126/science.adh2586

OECD. “OECD Principles on Artificial Intelligence.” *OECD Publishing*, 2019.

Otis, Nicholas, Rowan Clarke, Solene Delecourt, David Holtz, and Rembrand Koning. "The uneven impact of generative AI on entrepreneurial performance." *Harvard Business School*, no. 24-042 (2024): 84, http://dx.doi.org/10.2139/ssrn.4671369

Okunola, Abiodun, Adaan Ahsun, and Mike Blace. "Measuring the ROI of AI-Driven Workforce Transformation Initiatives." *Available at SSRN 5225996* (2025): 13, https://dx.doi.org/10.2139/ssrn.5225996

Patton, Michael Quinn. "Qualitative research & evaluation methods: Integrating theory and practice." *Sage*, 2015.

Peng, Sida, Eirini Kalliamvakou, Peter Cihon, and Mert Demirer. "The Impact of AI on Developer Productivity: Evidence from GitHub Copilot." *arXiv preprint arXiv:2302.06590* (2023): 19, https://doi.org/10.48550/arXiv.2302.06590

Rossi, Peter Henry, Mark W. Lipsey, and Gary T. Henry. “Evaluation: A systematic approach.” *Sage*, 2018.

Sauer, Carl René, and Peter Burggräf. "Hybrid intelligence–systematic approach and framework to determine the level of Human-AI collaboration for production management use cases."

*Production Engineering* 19, no. 3 (2024): 525-541, https://doi.org/10.1007/s11740-024-01326-7

Snyder, Hannah. "Literature review as a research methodology: An overview and guidelines." *Journal of Business Research* 104 (2019): 333-339.

Uwe, Flick. “An Introduction to Qualitative Research.” *SAGE Publications*, 2018.

Vaishnavi, Vijai K., and William Kuechler. “Design science research methods and patterns: Innovating information and communication technology.” *CRC Press,* 2015.

Van de Ven, Andrew H. “Engaged scholarship: A guide for organizational and social research.” *Oxford University Press*, 2007.

Weiss, Carol H. "Theory-based evaluation: past, present, and future." *New directions for evaluation* 76 (1997): 41-55.

Yang, Mohan, Nolan Lovett, Belle Li, and Zhen Hou. "Towards Dynamic Learner State: Orchestrating AI Agents and Workplace Performance via the Model Context Protocol" *Education Sciences* 15, no. 8 (2025): 1004, https://doi.org/10.3390/educsci15081004

Yin, Robert K. “Case Study Research and Applications: Design and Methods.” *SAGE Publications*, 2018.

**Appendix A. Interview Questions**

Participant Background:

1. Could you briefly describe your role?
2. What experience do you have with AI, automation, or managerial workflows?
3. Have you worked with decision-support tools or AI agents before?

Framework Discussion:

4. At first glance, does this model flow (Inputs → Processes → Outputs → Outcomes → ROI) seem realistic to you?
5. Are any layers unclear or missing?

Indicator List:

6. Which indicators across the five layers (inputs, processes, outputs, outcomes, ROI) do you find most relevant?
7. Which ones would be difficult or unrealistic to measure in practice?
8. Are there important indicators or effects that should be added?

ROI Model:

9. Does this hybrid structure reflect how your organization measures value?
10. Are the three categories (productivity, financial, adoption rate) clear and distinguishable to you?
11. Are any important categories missing from the formula?
12. How feasible would it be for an organization like yours to estimate:
    a. productivity benefits (e.g. time saved)?
    b. financial benefits (e.g. cost avoidance, margin effects)?

c. adoption rate (e.g. user trust, effect on its cognitive abilities)?

Challenges:

13. If this framework were applied in your organization, what practical challenges would appear?
14. Do you see any particular gaps or unintended consequences?

Overall Assessment:

15. On a scale from 1–10, how feasible is this framework for real organizational use?
16. What improvement would you recommend in order to implement the framework effectively?
17. Any final thoughts on the framework, indicators, or ROI model?

## Appendix B. Interview Transcripts

### Participant 1 (P1)

**Interviewer:** Could you briefly describe your role?

**P1:** I'm the managing partner at a small boutique consultancy firm in the realm of HR in talent management. I'm a co-owner and I manage daily operations.

**Interviewer:** What experience do you have with AI, automation, or any managerial workflows?

**P1:** I'm very curious about AI. This company started as a software development company 8 years ago. We had algorithm-based machine learning that would do a match between talents and opportunities in the market, and we called ourselves the tinder of recruitment. Right now, we are very much leaning to a more traditional executive search firm, and we also do the organizational development, but we don't do a lot of AI automation on that side of the business. Then, in my previous experience as an HR director, I worked with big multinational firms as HR director, and I've implemented some tools that were starting to play with AI and that automation back in the day, about 2018, 2020, 2021. Meanwhile, they were bought by Company X, but back in the day, there was a standalone company. It had a lot of AI both on a large language model and on predictive and analytical AI. So, I pretty much see myself as a pioneer in those things, but lately I haven't done a lot to be very honest. Hands on experience, I do teach and I'm the executive director of a course. Here, I gather a bunch of professionals that are spearheading this movement from companies that are really pushing the boundaries. So, I bring them together to form a cohort of facilitators and teachers that will bring their own experience and their own thoughts to the participants. I do it myself with my experience back in the day.

**Interviewer:** So, you are not currently working with any AI agents?

**P1:** No.

**Interviewer:** Alright, let's move on to the framework. At first glance, you saw that the model flow starts with inputs, then follows until ROI. Does this flow seem realistic to you?

**P1:** Yes, it makes sense.

**Interviewer:** Are any layers unclear or missing?

**P1:** On a very generic basis, I think it's very complete. Probably if we did a deep dive, we would find some bits and bobs, but generically it makes sense.

**Interviewer:** Let’s discuss the indicator list, so which across the five layers, inputs, processes, outputs, outcomes, and ROI, do you find most relevant?

**P1:** They are all equally important to be honest. Inputs define the quality of the data you put into the system. Processes define the efficiency or lack of efficiency if you have redundancies, non-adding value tasks within them. Outputs and outcomes, as well as ROI, are equally important. ROI will let you know if this is worthwhile or not, so all of them are equally important.

**Interviewer:** Which ones do you think would be difficult or unrealistic to measure in practice?

**P1:** Outcomes, broader organizational effects, you will need a lot of assumptions to narrow down the impact in other variables across the organization. So, the immediate test results, such as the amount of time you can reduce the amount of waste, are easy to measure. The outcomes you will need to factor in with a lot of assumptions. The ROI will have an impact on these, because you

will need the outcomes to measure it. Outcomes are difficult to measure, as a consequence, ROI will also be difficult to measure.

**Interviewer:** Are there any important indicators or effects that should be added?

**P1:** The notion of time, but I think it's pretty much within the concept of outputs, immediate task-level results. My hypothesis is that in outputs, you will measure anything related to time. Time element for me is really important because everything which has a cost in this life needs either energy or time to do it.

**Interviewer:** Let's discuss the ROI model. Does this hybrid structure reflect how you would imagine or how your organization measures value?

**P1:** Why do you call it a hybrid?

**Interviewer:** Because of the adoption rate element, it's the human-AI collaboration.

**P1:** Correct, it makes a lot of sense, thanks for clarifying that. If you have an amazing tool, but nobody uses it, it was actually a waste of money and time for everyone. So, the adoption rate makes sense, and therefore the hybrid approach makes sense.

**Interviewer:** Perfect, so all three categories are clear to you?

**P1:** Yes.

**Interviewer:** Are any important categories missing from the formula?

**P1:** So, the time is encapsulated within the productivity benefits in the form of time saved, shorter cycles, and reduced rework. If you don't want to go as mathematical, we could factor in a measure

of human engagement and motivation. It's not as objective as what you have here, but perhaps if you want to approach a different formula. You could combine both things together without impacting this one. I would be very pleased and motivated with my job if I could let go of everything that is time-consuming and non-value adding. Dealing with my agenda, dealing with booking meetings with clients or the team. If I let go of that, I would feel much more motivated. So, if we could capture that element of human motivation, all of a sudden, we could focus on what really matters. Then, it would be interesting to see those two elements combined, perhaps not in the same formula. One is very objective, other one is perception. As I'm thinking, it could even be the case that for some people, they would be upset because they feel comfortable. They find comfort in routine and admin, and transactional tasks. These effects also could be interesting to measure, because you get to know your people better.

**Interviewer:** I think that would be measured in the satisfaction that is in the adoption rate.

**P1:** Correct.

**Interviewer:** How feasible would it be for the organization to estimate the productivity benefits, the financial benefits, and adoption rate?

**P1:** I think productivity is quite easy. So, on a scale from 1 to 5, where 1 is non-feasible at all and 5 is very feasible. I think productivity is probably 5, particularly if we're talking about a streamlined process where you can measure time. Financial benefits are probably 3.5. My experience tells me that you always need some assumptions to calculate the financial benefits. For example, average time to do a task and average salary within the company, you combine both things together, and you get an average benefit in terms of financial benefits. Having said this, let’s make it 4 instead of 3.5. So, adoption rate, I would say also 5.

**Interviewer:** If this framework were applied in your organization, what practical challenges would appear?

**P1:** You'll need a lot of assumptions to measure the outcome section correctly. The other thing would probably be the estimation of all this, you have two points in time to apply this. One is before making the decision to buy an automation solution and after implementing it. Measuring what happened in the past is easier. When you want to estimate the ROI before the buying decision to justify if you want to buy it or not, then you will need a lot of assumptions of productivity gains, adoption rate, and so on. Then, in essence, it's the budget plan, an estimation exercise that may or not happen. So, that would be another trigger for me. How are we going to do this in a pre-active way, before making the decision?

**Interviewer:** Do you see any particular gaps or unintended consequences that this framework could have?

**P1:** Linking with the engagement thing I said earlier, it's basing your decision solely on a mathematical equation without having in consideration the human element. Sometimes it can backfire because you see ROI from a financial standpoint. You neglect the emotional and motivational aspect of the human being, either on one side or the other. This means people will either get upset if they find safety in doing things they used to. If that is the case, we need to seriously challenge that, but at least we are prepared. Alternatively, if you do this, there is much more benefit to unlock. Besides the process, you are making a lot of human beings much happier in their workplace. So that element could be interesting to capture.

**Interviewer:** On a scale from 1 to 10, how feasible is this framework for real organizational use?

**P1:** I would rate it 8

**Interviewer:** What improvement would you recommend in order to implement the framework effectively?

**P1:** Perhaps engagement again. If you don't want to overcomplicate, it might be just a simple question. Besides all the parameters in equation, just asking people on the scope from 1 to 10, how happy they would be by implementing automation. Measuring both things could be interesting to add.

**Interviewer:** Any final thoughts on the framework, indicators, or ROI model?

**P1:** Just outcomes, how do you measure coordination efficiency? Stakeholder satisfaction is a perception, but you can measure that. Capability development, knowledge of retention are complex concepts to measure. You might be able to measure, but it will be based on assumptions or perceptions. So, it's not as hard data as, for instance, outputs. Time saved per task, number of decisions, is very objective and concrete. So, there is a mix, and the outcome element is the only one I'm not giving you 9 or 10 in that scale, and I'm giving you 8. I think the whole thing is very consistent, well thought of. It's outcomes which might bring some difficulty into this.

**Interviewer:** Few questions about the impact assessment table. Are any important criteria missing?

**P1:** What I am missing here is money. Your model has the benefit of talking about money because it's ROI. Time is money, so you can always translate from one thing to the other. It's a very technical approach and functional approach, not a managerial one, in the sense that as a manager, you don't know the budget impact of AI. Error reduction, so there are two impacts of an error. One

is time because you need to rework something. The other one is money itself because you waste resources other than time.

**Interviewer:** How well can these four criteria measure the impact of AI?

**P1:** I like yours better because it has the adoption rate and money impact, which we are missing here. Although you could deduct it from time and errors.

**Interviewer:** Okay, perfect. Thank you.

**Participant 2 (P2)**

**Interviewer:** Could you briefly describe your role?

**P2:** Right now, I'm the chairman of 2 companies and I'm a non-executive director of another one. So, I no longer have executive roles, even though in the past, I've been in charge of planning and controlling a new business of a conglomerate for 12 years. I was a vice CEO of a bank, with responsibility for internal processes and planning, so organization of the bank. I've done other things, but these were the most important roles I had for a long period.

**Interviewer:** What experience do you have with AI, automation, or managerial workflows?

**P2:** Managerial workflows in all those cases. When I was in banking about 15 years ago, my view was that usually bank managers did not pay enough attention to the organization and corporate functions, and how to optimize processes. So, I spent a lot of time doing that. At the time there was no AI, but we already had robots and lots of automation. It was a very large retail bank, with branches in different countries on different continents. Automation and standardization were the main concerns in the source of productivity increases, and compliance helped because I was not subject to human error in some procedures.

**Interviewer:** Within these automations and managerial workflows that you were working with, did you have specific decision support tools, you mentioned you haven't used AI?

**P2:** At the time no, but as a consultant in the late 90s, together with some colleagues from the engineering school, we created routines for a bank and powered some of their analytical tools using neural networks. Today would be called AI, but I didn't follow through working on that. Also, because my AI colleagues, the engineers, turned out to do other things. We were in a way ahead

of our time, but not true enough to understand that there was a path for the future. This was in the 90s, and it was basically to help them in the credit process.

**Interviewer:** Very interesting, especially when you work with something where you cannot describe it yet or not give it a name yet.

**P2:** Yes, but at the time we just used neural networks, we explained how they worked. Then, we showed that it would have excellent fits, better than econometric fits that I also had to compare. So, they were fairly convinced, and they used them for quite some time.

**Interviewer:** Let's now step into the framework. At your first glance, does the model flow we are using going from inputs to ROI seem realistic to you?

**P2:** Yes, it does for standard processes that it aims to describe. There may be some more complicated processes where you would need something different, but for a lot of processes I know and I've worked with, it's an insightful tool.

**Interviewer:** Would you say there are any layers which are unclear or missing?

**P2:** Not that I'm aware of.

**Interviewer:** Next, we will discuss the indicator list. Which indicators across these five players do you think are the most relevant ones?

**P2:** I would highlight the second one [integration and development] and the data set because there's more money involved in the hardware. The point is, for the most part, it's not so difficult and it's usually something where I rely on expert opinions.

**P2:** Well, the bottlenecks are always key.

**P2:** Decision cycle time.

**P2:** I would say consistency and time saved per task.

**P2:** [Outcomes] are also very relevant. I tend to overlook one that is key, which is other stakeholders, because if it worked for me and the people that were dealing with me, I would be satisfied. You have to also take into consideration the interaction, how it may help or harm other processes not involved in this specific automation or robotization. So, it's important to consider all of them [stakeholders], all involved and affected.

**Interviewer:** The last category we are looking at is the ROI, here we have five points we highlighted, the productivity benefits, the quality benefits, the financial benefit, and finally the total cost of ownership.

**P2:** The last two.

**Interviewer:** Would you say some are difficult or unrealistic to measure in practice?

**P2:** No, you have to work out some of those ideas in measurable metrics. You need a specific metric, but they're all measurable. For example, you have to calculate total cost well because a lot of times it may not be total. There's always a metric associated with or at least an approximation.

**Interviewer:** Would you add any effects or indicators that you have not heard yet?

**P2:** Well, there's one that is always present, but it's very hard to measure, which is the opportunity cost. Once you automate, you can create a lot of value doing other things.

**Interviewer:** Next, we will talk about the ROI model. Does this hybrid structure reflect how the organization measures value in your opinion?

**P2:** Yes, I think that it covers the main relevant areas.

**Interviewer:** And are the three categories clear to you?

**P2:** Yes, very clear.

**Interviewer:** Are any important categories missing from the formula?

**P2:** There is no obvious category missing, I would have to think harder on that. As far as I understand, I think it's comprehensive.

**Interviewer:** How feasible would it be for an organization to estimate productivity benefits, do you think?

**P2:** Well, it depends. We still have organizations in 2025 that do not have internal analytics in terms of gathering all the information required and necessary to handle that. In the past, I've been involved in formatting SAP processes for companies. Usually, the information that we talk about is available in the company. You can extract it as long as you design the information mechanism. From the beginning, the managers are willing to get all the information that at some point may be important for decision making.

**Interviewer:** So, would you say that your answer refers to how feasible you think you would be able to measure productivity benefits, financial benefits, and adoption?

**P2:** It depends on the information you get currently. I've been in companies where I had information about every single process. In other companies, you only have aggregate data, then you have a hard time measuring productivity. So, it depends a lot on how you lay out your information system, regardless of using this methodology or others.

**Interviewer:** Of course, it always depends on the industry and so on.

**P2:** Yes, and the practice of management, I will give you an example. In 2007-2008, I was not directly managing that company I've been to before, but the holding company of a chemical group that still exists and is growing. For example, we had a plant and automated all the supply and expedition processes. We could do that because we had instant information about the cost of each and any of the options. So, if I cannot automate and don't have the information, then I may not be minimizing cost or redundancy.

**Interviewer:** What practical challenges do you think would appear if this particular framework was applied in an organization?

**P2:** I think the main challenge is on-time micro information to make it work. The problem is do we or don't we have the information? So, how desegregated and micro the information is within the company that is going to determine whether you can have a good instrument or something too broad, and therefore not so useful.

**Interviewer:** Do you see any other particular gaps or unintended consequences of the framework?

**P2:** Well, in the end you have some human choice in using that. So, I think the problem is always whether the person or the group of people deciding is taking the right advantage of the instrument, I don't see any negative consequences of using it.

**Interviewer:** Overall, on a scale from 1 to 10, how feasible would you say this framework is for real organizational use?

**P2:** For more sophisticated companies, I would say 9. You would have a lower value if the management were not information-based. A lot of people manage companies only with aggregate information, they do not come deep into the processes regarding automation or AI. You can make better decisions if you have very detailed and on-time information. So, with the state of information systems, I would say 9. Just not to say 10 because I think it's totally feasible.

**Interviewer:** What improvement would you recommend in order to implement this framework effectively?

**P2:** Well, in practice when you start applying, you may not be able to do exactly as it is said in the framework. So, you have some shortcuts, if any of those things that I think would be available is not available. So, we always need to have the possibility of generating alternatives that are coherent within the system. It's like a warning, you cannot foresee them in advance, but you have to let the system function even if you're not at the optimal level at each of the variables.

**Interviewer:** Quick question about the impact assessment of the AI tools. Maybe you've heard of the four criteria, which is the time saved from implementing AI, also its potential to reduce errors, its scalability impact and the value of the time that is saved to allocate it to more strategic value tasks. Do you think that any important criteria missing?

**P2:** No, because your last one, it's a catch all. It provides the opportunity cost, which is the best indicator of the utility of the process. Well, you can go for more detailed things, workforce or input reduction, but you already have that when you look at cost efficiency.

**Interviewer:** So overall, you would say that all of them can measure impact well, right?

**P2:** Yes, especially if you put them together.

**Interviewer:** Do you have any final thoughts on the framework, indicators, or the model itself?

**P2:** It's to be well applied, very information sensitive and demanding. If you want to take advantage of AI, we have to have data.

**Interviewer:** These were all the questions. Thank you.

**Participant 3 (P3)**

**Interviewer:** Could you briefly describe your role?

**P3:** I am the CEO of a company dedicated to digital transformation using AI as well as enterprise project management, also using agents these days. My role involves sales, high level discussions with clients on what can AI bring to them.

**Interviewer:** What particular experience do you have with AI, automation, or managerial workflows?

**P3:** We do a lot of agents for organizations across the company. We do process automation by using agents in Microsoft technologies. So, we've done a lot of agents over almost 2 years since Microsoft launched the agent framework. We were already using agents but in a different way, we used AI automation scenarios and agents for workloads.

**Interviewer:** Before that, have you also worked with any AI agents or decision support tools?

**P3:** Before that, we used a source OpenAI when Microsoft launched it. That was about 2.5 years or almost 3 years ago. So, we've been using agents always in an enterprise scenario, not in a consumer scenario.

**Interviewer:** Let's move on to the model flow from inputs to ROI. At first glance, does it seem realistic to you?

**P3:** I don't see it as much as a mathematical equation because a lot of things here will have an impact on this flow. Although, when we're doing a business case, we need to understand the expectations from an investment perspective. When we get down to business, a lot of the impact

of AI boils down to the human factor. If people are not open to using AI in their processes, that will have a huge impact on the ROI. So, it depends on how people work with or work against AI. We see a lot of that in the results that we do. Going back to here, if I understand the formula correct but you're here to help me out.

**Interviewer:** For now we're just discussing the flow itself, not the calculation model.

**P3:** Well, you need to explain processes. How does that translate into a numeric variable? Is there a number behind each of them?

**Interviewer:** Not yet, we are validating whether the framework is clear, practically applicable, and relevant.

**P3:** I'm not sure if I would put here the type of process because it will be interpreted by the number of hours being saved. So, it's not because if it's autonomous, it's going to save more hours ahead. It still requires some control in there, but it's going to save some hours.

**Interviewer:** If I understood you correctly, it seems realistic, but you are unsure about the processes?

**P3:** Yes, I think the formula is important. For instance, I would put a variable that is related to the maturity level of the organization. If it's an organization that has people that have been trained, they are willing to work with AI, then the output will have more impact than giving them something to do.

**Interviewer:** Would you say there are any other layers missing or unclear in this flow?

**P3:** Well, the adoption rate might interpret the maturity level. Trust is important absolutely. The other one is the willingness to actually work that is missing, but it's part of the maturity level of the organization.

**Interviewer:** What about the layers in the model flow, so the inputs and processes? Would you refer to what you said earlier, maturity?

**P3:** Yes, I would add it in there.

**Interviewer:** If we now look at the hybrid structure of the formula itself, would you say that this reflects how your organization measures value?

**P3:** What do you call as an A within ROI model?

**Interviewer:** This particular formula is hybrid because it has the adoption rate, human context, element that reflects human-AI collaboration. So, there are benefits that AI brings into organization, primarily productivity and financial ones. Then, there's an adoption rate, so how people react and how all of this combines overall to understand the impact that AI has. It might be that it brings benefits, but then people are so upset with, for example, not being able to do their routine tasks. Then, the ROI decreases much more because of this component. So, if we balance it together, then the overall impact can be measured more precisely.

**P3:** Okay, so it's looking into the human factor as well as the financial factor. So, adoption is very important, and I think you need to go a little bit further. It's not just user satisfaction, I would say that maturity needs to be translated into human capability of using AI for non-autonomous models. I think you need to add the capability of the person in terms of training and willingness to actually use and trust the system. For the total cost of ownership, you mentioned opex, operational

expenditures of using the system, but what about implementing it? Designing, buying the services to implement AI because a lot of the time, people don't know how to create their own agents. A lot of the agents, if they are simple to implement, they'll probably have very little benefits. Typically, you need some specialized services to connect them between systems, called APIs. So, the total cost should be capex and opex, So, initial cost that you have to implement, plus the ongoing usage of the model.

**Interviewer:** How feasible would you say it is to measure other elements?

**P3:** I think they're meant to measure. If you want to put them into formula, you need to convert all of that into cost or value. So, you need to understand the cost of having people in there. Let's say, if one of the outcomes is, you don't need 3 people, you only need 2 people, because now you're going to have a more productive system. Depending on the country where you are, removing those 2 people will have significant cash out in terms of severance fees. So, we need to pay the people to live, pay an administration, which could typically amount to the number of years that the people are in the organization times 1.5 salaries or 2 salaries. So that is also something that you need to question when you're implementing AI, which is the cost of severance fees to people. You can also expect that because if you're just looking at the productivity, you say, I'm not going to pay this salary, but I have to pay severance package here for them. If people are in the organization for 20, 30 years, the severance package for them before they reach retirement could be quite significant. Typically, companies are willing to pay for that because the total cost of having the salaries outweighs the severance package, but it depends on the side where people are. It's a difficult calculation when we're talking about people times productivity cycles. Of course, if you are qualifying people that instead of them doing this, now they're going to do something else, they are not going to waste this time. If you don't have anything else for people to do, that also brings that

value, just saying that the cost of these people will continue to be in the organization, but they're not going to be doing this. You cannot just say that I have savings here or a fixed amount of hours times the salary of that person. In reality, you're still going to pay the salary of that person anyway. You might say that by putting the person doing something else, we will offset that. That's fine, but the calculation behind this is extremely complex. There's a ton of variables on how to calculate these things.

**Interviewer:** So, would you call this a practical challenge when applying this framework to an organization?

**P3:** Yes, because for instance, the company like ours, we have to deal with a lot of calculations that Microsoft, Google and all the others pulled out. They say, you're going to have savings of 30%, 40%, and so on. Then, when my clients ask me, how do they come to this figure of 30%, I say I'm not Microsoft, so I have no idea how to come to that figure. When I ask Microsoft, they have done calculations, they or anyone else in the market. They are typically done by some consultants that they hire, sometimes McKenzie or other guys they will hire to calculate something. The calculation is always very high level, and it's done by people who are not really into the practicality of these into organizations. It just makes some very big figures on this. So, when we go into the clients, we are the ones who have to dive deep into these things. We tell the client how much this is going to save and cost for this particular process. Microsoft doesn't know that, or Google, or OpenAI, they don't know very high numbers that will escalate their value. Then, we need to look at the reality of each organization. The same process in different companies might have different savings.

**Interview:** On a scale from 1 to 10, how would you rate this framework for real use?

**P3:** I would have to put it at 4 at this point because for the cost you need to have capex. I don't know how you're going to calculate productivity and financial benefits in a very easy way. I think you guys need to dive a little bit into that. The adoption rate is a very good one, but I don't see how we're going to put that into an index because it might depend on the type of department. Let me ask you something, are you going to calculate this ROI for an entire organization in adopting AI or process by process?

**Interviewer:** So, adopting AI in overall organization.

**P3:** That's going to be very difficult because if you implement AI for a process within IT, it's one thing that your adoption variable is going to be different than if you're going to implement this in a department where people are close to 60 years old. They are not very familiar with or fluent in IT, it's going to be a bit more complicated on adoption. So, if you want to implement adoption rate for the entire organization, then definitely 4. If you want to implement this process by process, I think we can maybe push to 5 or 6. You always need to have capex in there. You need to have a lot of people just implementing and designing AI in a proper way, external people, specialists to tie up between the interfaces and the systems. It's not so easy. If you're talking just about generative AI to make a text or to implement to facilitate the way you write your email, that's one thing. If you're going to use AI as agents that's a very different thing. It brings a lot more benefits, but you need to have much more specialists because they need to connect to your systems.

**Interviewer:** Which indicators across the five would you say are most relevant and which ones would be difficult to measure, maybe there are some that should be added?

**P3:** Ah, there you have integration and deployment. So, let me give you another point, that's one of the big costs here. So, licensing and maintenance cost, integration and deployment, it's relatively

easy. Data preparation, I don't think you need to prepare anything almost, that's the beauty of AI because unlike previous data models, the system can work very well with abstract data, so that's greatly reduced. Typically, we use cloud systems instead of IT infrastructure.

**Interviewer:** It depends on the industry and the type of AI they incorporate.

**P3:** Yes, unless you want to run an AI locally. Even if you want to use the proprietary AI, you typically use cloud, not hardware. That's very uncommon, probably mainly in medical and pharmaceutical. I think you also need to put people from your side explaining the process for the people who are going to implement the AI. So, redesigning the processes is also an important cost because it will require a lot of senior people from your side to explain what the output is that you want, not necessarily how you're doing it today.

**Interviewer:** So, would you say that the inputs are most relevant in this case, or would you say that all of them are?

**P3:** Yes, I would say that your time to rethink the processes internally together with a specialist organization, that's a significant amount of time as well. It doesn't only cost you hours, but it will also cost you focus on what you're doing, selling in the organization. Then around processes, autonomy ratio, I wouldn’t go here because this is already translated in the variables for you to calculate the benefits. On the inputs, you also need to add the cost of the resources that are required today to produce the process. Maybe you should add in there if the process is more generative or classic, which is, does it require reasoning or does it require just an interpretation, or it is just a workflow where you need to do some calculations. Depending on that, it might yield different benefits.

**Interviewer:** So, now that you've had a look at this, would you then change your score on the framework itself?

**P3:** Yes, I would put it as 5 if you're doing it for the entire organization, or as a 6/7 if you're doing just for one process.

**P3:** Time saved per task, I don't think that's really an output because you need to convert that into a cost. So, time will need to be converted to cost, unless it is an intermediate output for them to calculate. Throughput depends on the industry, the rest seems okay.

**Interviewer:** Do you have any final thoughts on everything that we've discussed?

**P3:** Overall thoughts, I don't think you should use ROI, sorry about that. I know it's a big thing for you. ROI is a bit flat for that analysis, but it's a place to start. It brings a lot of simplification, but I really think that it's interesting when these benefits are kicking in, that's important. Productivity benefits, financial adoption rate, total cost of ownership are variables that will impact this. Bear in mind that productivity benefits, everything translates into financial benefits, and the total cost of onwership needs to include both capex and opex.

**Interviewer:** A final question on the impact assessment table. How well can these four criteria measure the impact and are any important ones missing?

**P3:** No, I think this is good and adds a lot of value to the model. Just one thing, instead of strategic value, I would just use value added activities or higher value generated. The term strategy is used in many things in an organization and not everything is strategic. This is in my market when I talk with another CX level in the organization, we avoid to overuse these terminology. Strategy is when we are making decisions on where the company's should be positioned, areas, markets, offering,

and so on. When we're dealing with performing higher value activities or doing things better, we are typically talking about tactical levels. So more things like a higher value, generating more value in what we're doing in the day-to-day. If it's to do different things, then it's strategic. If it's to do the same things, but with better value, then it's tactical levels. Apart from it, I think from a model perspective, it looks good.

**Interviewer:** Okay, thank you, that was it.

**Participant 4 (P4)**

**Interviewer:** Could you briefly tell us about your experience, your professional role?

**P4:** I'm an associate professor for 7 years, so faculty people who specialize on the teaching site, not research, so I bring things and content to students. I've been focusing mostly on innovation, product design, and development. I'm an engineer by training, computer scientist. So digital technology is my comfort zone. I play around with these things and try to figure out how they can be useful for companies or people who are in already a working environment.

**Interviewer:** Do you already have some experience with AI, automation, or managerial workflows specifically?

**P4:** Yes, a little bit but not directly because I've been recommending companies, so more of a consultant side. So, I sat on the board of a media company for a few years, and this is one of the things that we discussed.

**Interviewer:** Also, in your private life, have you worked with decision support tools or AI agents before?

**P4:** Yes.

**Interviewer:** At the first glance, does this model flow from input to process, output, outcome, to ROI, seem realistic to you?

**P4:** I'm missing the measurements of the ROI. The productivity benefits are not something that only reduces the energy put into the system, so you get the same output at the end. It's a cheaper

version of running that process, but what if you can actually get a better outcome than you previously were able to. Other than that, so this first part went by perfectly fine.

**Interviewer:** So, no unclear or missing layers that we could integrate?

**P4:** No.

**Interviewer:** Next we will go to the indicator list. Do you see any indicators that are either the most important, or difficult, or unrealistic to measure in practice?

**P4:** So, where will you include the actual engagement of when people are operating these tools. If company doesn't have anything, we should invest in these agents. So, when that decision comes, then you'll look, do we have the software or do we need a license. Then, what are they going to be working with, so do we have data preparation. For instance, do you have all the information that's ready to be accessible. Is it codified in a way that it's regular, uniform and can be used. Oversight is monitoring, so alarms and anything that needs to be reset. Training is preparing people to be able to use these. To build their own agents to do the different things that they think will be useful to them, so they can leverage that. What about the day to day, the regular operation, where's that?

**Interviewer:** That's in the process. Also, training could be both, not only the people who are being trained to use AI agents, but also the AI to be trained with the right proms and new data to grow. The process is where people use them and where after activation of the AI agents in the daily business, we do the error tracking, the escalation rate, the repetition rate. While it's working, we then track if there are errors bottlenecked, what the whole cycle is, if the decision cycle makes sense, if we can connect some AI agents, if they can also work together in many times. So, the actual working of AI agents is in the process.

**P4:** Okay, good.

**Interviewer:** For each process or layer, do you have one indicator where you say this is the most relevant, or would you say they are all relevant?

**P4:** They're all fundamental. So, if any one of these is zero, this is a product of all of these, then everything is broken. So, you have to have them all. If there's a more important one, I'm going to go to an engineering answer, which is, it depends. Here's my thought process. If you have an IT infrastructure that's robust and very capable, then what's going to matter most is if you have the best models to run. The flip side of that is I don't have a good infrastructure. So, I can only use one, then the IT infrastructure becomes the most important thing because you have to invest in hardware. So, maybe there's precedence, you need to have the right hardware to figure out the right model. If you have the best model, then you could build an infrastructure that's not over budget to keep the demands of that. So, it's a push pull thing, it's an intention, but you need all of it.

**Interviewer:** The next question would be, are there any indicators where you would say these ones are very hard and unrealistic to measure in real life?

**P4:** It depends on what you want, so if the outcome of the output is something that is a quality of nature or probably not. So, how repeatable is the process. That's usually a thing that you use to compare it. So, I'm going to use that analogy to try to explain what I'm saying. If you have a factory that's manufacturing the same cup over and over again, then you can actually measure it. On the other hand, problem with optimizing the design process is that we have a challenge. We have one client, and you invent something, you can't reinvent the same thing afterwards, because you already came up with it. So how do you know that thing was the best one? Whereas in the other process,

you had a repetition, so you could actually measure and see if it was better or not. So, accuracy in that case is going to be very hard to measure.

**Interviewer:** Lastly, are there any indicators that you might not have seen or that you're missing in the list?

**P4:** Increase of quality. It seems to be that you're just optimizing for the same thing. For example, we're still manufacturing cups, but we're using less energy. How can we actually formulate that we can go much further because now we have these tools? So, it's an improvement or an expansion. So, what I'm looking for is not just that it works, but it works better. Let me give you another example with self-driving cars. So, people are working on technology, so that cars could drive themselves without having a driver. The threshold is, can we get cars to drive as well as humans? Which is a mistake already in itself, because if you look at the statistics, there are tons of accidents all the time. So, actually cars cannot have already been able to drive better than humans for maybe 10 years already. The threshold that we put on them is that they can't have accidents, so that's what we expect. What if our level expectation would actually be that they don't need to drive with humans, but they could drive at the level that humans can't. So, we expect cars to be able to drive alongside humans, but these systems can communicate and coordinate. Using the stoplight example, there's a circus number where they're riding on bicycles, and then they start intersecting, and it looks like they are going to crash because they go between the gaps. Imagine an intersection where people would try to do that, guaranteed accident, but cars could do that. So, this is improved efficiency throughput because they could coordinate much faster, thousands of times per second. So, this is something that we could aspire to, we can have an even more efficient thing that humans can't do. That was my idea here.

**Interviewer:** The next topic is our ROI model. Do you think that this hybrid structure does reflect how an organization or maybe an organization that you consulted and worked for measures value?

**P4:** I don't think anyone actually uses such a theoretical understanding of this. Management ends up being a more practical decision-making process of using some of these factors as arguments to leverage the decision.

**Interviewer:** Do you think that these three categories you are seeing, the productivity, financial adoption rate, clear and distinguishable?

**P4:** They are clear.

**Interviewer:** For the missing part, you've already mentioned the quality measurement.

**P4:** Yes, adding opportunity for an increased quality, so what we could do with this that we could not do before. So, I think there might be one that's temporal, that maybe I would include it today but in 10 years time, it's probably going to be irrelevant. It's reputation, so what would happen if you let some of AI agents by themselves and they screw up. So Microsoft had one of these chatbots 3 or 4 years ago to interact with people. Within one day or something, it became racist, so there was a reputational damage to the company. So, what are the guarantees that you know this will not affect the reputation, the brand, the quality of the organization. Again, it depends on what you're doing. It might be that if it is for industrial setting, this thing screws up, doesn't work. At the end of manufacturing, you see that this product has a defect, so it's just a cost. If it is a product that ends up being used by people and then it explodes. Samsung a few years ago had an issue with one of their batteries because of the way they were manufactured. They were trying to make it so thin that a part at the corner of it when you dropped your phone would short circuit, so the battery

would explode. So, we are talking about a moment in time when they failed. There's also a lot of novelty here, we are not entirely sure what are the consequences of doing these things.

**Interviewer:** Next, the potential risks. If this framework were applied to one of the organizations that you are working with, what would be the practical challenges that would appear and any consequences or gaps that you could see? Do you think that there is some obvious risk in putting this framework into practice?

**P4:** I think the risk is always there when deciding to make an investment. If there were no risk, the decision would already have been made. I think the risk here is more of how long or how mandatory these investments are that need to be made, and how the company is going to decide. So, your framework helps an organization make sure there's nothing missing in the analysis. You need all of this, so using this reduces the risk of companies being sure or being more convinced that they're making the right decision.

**P4:** So, the tool itself minimizes risk, but then we could also be pessimistic and say, now I have a great tool that can help me. So, I'm going to be more convinced to take the risks of looking into this and we adopt it because we use the tool, but it’s a pessimist way of looking at things.

**Interviewer:** Finally, as an overall assessment, from a scale from 1 to 10, how feasible is this framework in real life?

**P4:** I would say there's nothing preventing people from using it. If they decide to, so they can. So the feasibility on that metric, I think would be 10. It's absolutely very easily adoptable. Will they though? That goes back to a previous comment that I don't think that organizations actually are disciplined in their decision-making to go through the complete process that the care that you put

into making this. So, maybe there's the feasibility, the usability then is maybe a 5 because if they want to, they can, I just think that they don't.

**Interviewer:** That's a big difference, obviously. That was the final question. Thank you.