

Alvertos, Konstantinos; Ioannou, Nikos; Kokkinis, Dimitris; Katsianis, Dimitris;  
Varoutas, Dimitris

## Conference Paper

# A Decision-Making Framework for Integrating Generative AI in Enterprise Workflows: Building vs. Buying

ITS 33rd European Conference 2025: "Digital innovation and transformation in uncertain times", Edinburgh, UK, 29th June – 1st July 2025

## Provided in Cooperation with:

International Telecommunications Society (ITS)

*Suggested Citation:* Alvertos, Konstantinos; Ioannou, Nikos; Kokkinis, Dimitris; Katsianis, Dimitris; Varoutas, Dimitris (2025) : A Decision-Making Framework for Integrating Generative AI in Enterprise Workflows: Building vs. Buying, ITS 33rd European Conference 2025: "Digital innovation and transformation in uncertain times", Edinburgh, UK, 29th June – 1st July 2025, International Telecommunications Society (ITS), Calgary

This Version is available at:

<https://hdl.handle.net/10419/331248>

## Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

## Terms of use:

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# A Decision-Making Framework for Integrating Generative AI in Enterprise Workflows: Building vs. Buying

Konstantinos Alvertos<sup>1</sup>, Nikos Ioannou<sup>1</sup>, Dimitris Kokkinis<sup>1</sup>, Dimitris Katsianis<sup>2</sup> and Dimitris Varoutas<sup>1,3</sup>

## Abstract

The use of Generative AI (GenAI) in recent years has significantly reshaped workflows across multiple industries, enhancing productivity and competitiveness. Enterprises commonly face the strategic decision to either build proprietary GenAI solutions or buy these services from established vendors. This paper proposes a techno-economic quantitative decision-making framework designed to clarify the “build vs buy” dilemma by quantifying the value derived from GenAI implementations, based on the employee productivity increase. We introduce a model that evaluates the break-even point, the specific number of employees whose cumulative productivity gains justify transitioning from purchasing external GenAI services to building in-house solutions. Our analysis indicates that this threshold varies distinctly by departmental functions and organizational size. Ultimately, this research aims to equip enterprises with actionable insights to inform strategic decisions about GenAI investments.

## Keywords

Artificial Intelligence, automation, productivity, technology adoption, investment, strategy

## 1 Introduction

The rapid evolution of AI, particularly GenAI, has introduced a transformative era for workplace productivity. Empirical evidence has shown the capabilities of GenAI to increase productivity and solve complex, knowledge-intensive tasks within enterprises. Recent empirical work has demonstrated the potential of GenAI to enhance productivity and streamline complex, knowledge-intensive tasks within enterprises. As shown in [1], experimental evidence indicates that GenAI adoption can significantly improve task completion times and output quality in knowledge-intensive activities. Building upon these productivity findings, [2] presents field experimental evidence specifically within the e-commerce sector, showing that GenAI-driven enhancements in customer service workflows can yield measurable performance improvements for workers. Beyond immediate operational gains, the economic consequences of AI integration extend into innovation and long-term growth trajectories, as shown in [3], which underscores the role of advanced AI tools in stimulating productivity at the research and development level. At a more strategic level, firms not only face the question of whether and how to adopt AI. The analysis in [4] examines the “make or buy” decisions for AI technologies across multiple industrial contexts, highlighting how sectoral characteristics influence optimal sourcing strategies. Moreover, decisions to implement AI-based solutions also require a keen understanding of the cost-effectiveness of automating specific tasks [5], and AI’s macroeconomic implications [6].

---

<sup>1</sup> Dept. of Informatics and Telecommunications, School of Science, National and Kapodistrian University of Athens, Athens, Greece

<sup>2</sup> Dept. of Digital Industry Technologies, School of Science, National and Kapodistrian University of Athens, Psachna, Greece

<sup>3</sup> Corresponding Author: [d.varoutas@di.uoa.gr](mailto:d.varoutas@di.uoa.gr)

Despite these advances in understanding the productivity and economic implications of GenAI, the current literature does not offer a comprehensive framework to guide enterprises in quantitatively assessing the financial and strategic trade-offs of adopting GenAI. Specifically, no existing model systematically integrates both the total cost (capital and operational expenditures) and the multifaceted value, such as productivity increments, quality improvements, and strategic advantages, that GenAI can bring. Without such a structured framework, enterprises lack a clear, evidence-based method for determining whether acquiring a GenAI Software-as-a-Service (SaaS) product or developing a proprietary GenAI model would yield greater long-term returns.

The following research aims to bridge this gap by introducing a novel economic framework that calculates and compares both the costs and the derived value of GenAI adoption under different scenarios. Additionally, by leveraging insights from the literature, the framework calculates the “pivoting point” at which the economic and value proposition of building a custom GenAI model surpasses that of purchasing an off-the-shelf solution.

The remainder of this paper is organized as follows: Section 2 presents the proposed techno-economic framework and the methodology to capture the decision threshold. Section 3 presents an example application of this framework, providing quantitative results regarding the costs and value of GenAI adoption across different enterprise scenarios. Finally, Section 4 discusses the implications of our findings and concludes the research with recommendations for enterprise decision-making regarding GenAI adoption.

## 2 Methodology

The proposed methodology considers two different scenarios: build and buy. A build scenario comprises an open-source model that an enterprise trains on proprietary data and deploys within its own infrastructure (on-premises or cloud environments). The buy scenario refers to an AI model that is vendor-supplied and pretrained by the vendor to perform specific tasks via managed APIs or platforms. The goal was to calculate the lowest user threshold for an enterprise to transition from buying an AI model to building its own.

For each of these scenarios, we calculate three dimensions associated with economic performance:

- 1) The value of GenAI to the enterprise, expressed as productivity gains for employees.
- 2) The initial capital expenditures (CAPEX)
- 3) The ongoing operational expenditure (OPEX)

By combining these parameters and applying a structured analysis that examines the response of the model to incremental changes, using the number of employees as the variable we can select the more cost-efficient scenario for an enterprise. An overview of the methodology is illustrated in Figure 1.

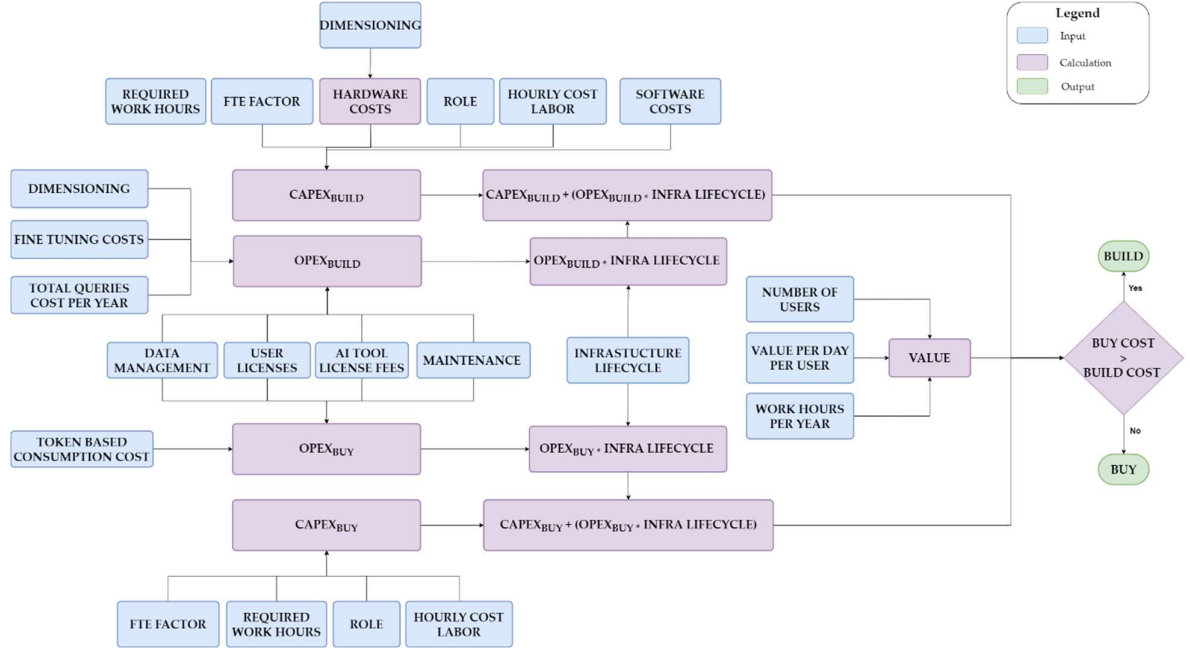


Figure 1: Overview of the methodology.

The use cases under consideration (i.e., the enterprise procedures that will show productivity gains when GenAI is applied) are illustrated in Figure 2: Computer Programming, Business Productivity, Customer Support, Marketing and Document Search and Summarization.

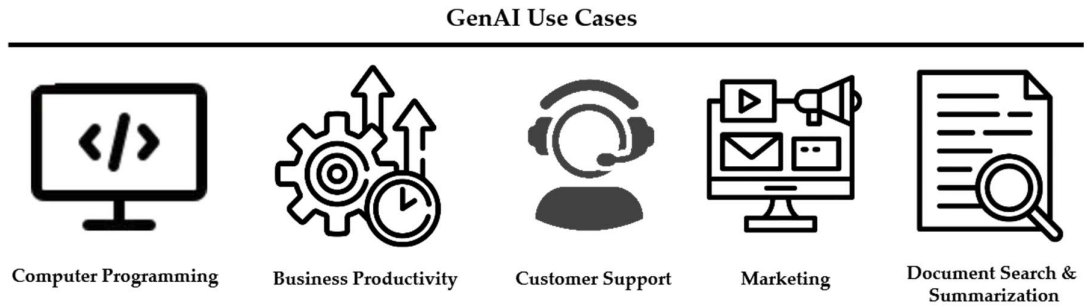


Figure 2: The use cases evaluated for GenAI

The integration of GenAI in these contexts yields numerous benefits.

In the domain of computer programming, there is a 6.5% increase in successful code contributions and routine coding tasks, accompanied by a 55% reduction in task completion time, without any significant compromise in code quality. However, this is offset by a 41.6% increase in integration time due to heightened coordination costs [7].

In the business productivity sector, GenAI facilitates skill development, onboarding processes, and problem-solving training, thereby enabling employees to concentrate on complex tasks. Task automation enhances operational efficiency, leading to cost reductions and streamlined workflows. Furthermore, GenAI supports finance and logistics by augmenting supply chain management and fraud detection [8].

In customer support, AI agents can simulate human interactions, reducing the average response time during customer support calls from 8 hours to 9 seconds (99.97%) and providing 24/7 availability. Post-implementation surveys indicate that 85% of clients who interacted with an AI agent found it beneficial [9]. Nonetheless, challenges persist, particularly with limitations in NLP technologies, where multi-part queries remain unresolved.

In marketing, GenAI can generate personalized storytelling content and images, resulting in increased sales and brand loyalty. Targeted storytelling aligned with customer journeys and personalized narratives enhances sales funnel efficiency and provides insights into hidden consumer preferences through data patterns [10].

In document search and summarization, where human review of legal invoicing requires 194 to 316 seconds, a GenAI system requires only 3.6 to 21.5 seconds per invoice [11]. In administrative and governmental applications, automating government reports with GenAI reduces manual report revision activities by 80-90%. However, limitations arise when visualizations lack explicit data labels. These findings underscore the potential for scalable deployment in reporting workflows when hybrid human-AI validation is employed [12].

## 2.1 Value Calculation

The value created from GenAI is derived from the time saved per day by the user within the enterprise.

Let  $T_{SAVED}$  represent the potential time saved per day, calculated as

$$T_{SAVED} = TT_{AVG} \times P_{GAINS} \times (1 - P_{LEAK}) \times H \quad (1)$$

Where  $TT_{AVG}$  is the average time spent on a specific task (% of total hours),  $P_{GAINS}$  the percentage of productivity gains (which differ depending on the task),  $H$  the number of hours within a workday and  $P_{LEAK}$  the percentage of productivity leak, i.e. proportion of productivity gains lost since a user can spend some of the time saved non-productively .

Let  $V_{PER DAY PER USER}$  denote the value created by GenAI within a working day for each employee per day per user, calculated as

$$V_{PER DAY PER USER} = HL_{USER} \times T_{SAVED} \quad (2)$$

Where  $HL_{USER}$  is the hourly labor cost per employee and  $T_{SAVED}$  is the time saved as expressed in (1).

Finally, value gained for an enterprise can be defined as

$$V = U \times V_{PER DAY PER USER} \times M \times D \times H \quad (3)$$

Where  $U$  is the number of employees,  $V_{PER DAY PER USER}$  calculated in (2),  $M$  is the number of months within a calendar year, and  $D$  is the number of active workdays within a month.

## 2.2 Initial Costs of GenAI (CAPEX Calculation)

For both scenarios, the initial cost (CAPEX) refers to the establishment of the GenAI platform, encompassing all upfront investments required to acquire, install, and configure the necessary hardware, software, and the supporting infrastructure. This includes purchasing servers, GPUs, networking equipment, data storage systems, and any facility upgrades required to support high-performance computing. Additionally, CAPEX includes the one-time costs for software licenses, integration, and initial setup services, all of which are necessary for the GenAI platform to become operational.

In the build scenario, we need to calculate the runtime throughput, GPU sizing, hardware and software expenses, and the necessary labor costs.

Let  $Q_{PS}$  represent the total queries per second calculated as

$$Q_{PS} = \frac{U \times Q}{60} \quad (4)$$

Where  $Q$  are the average queries, a user is expected to input within a minute.

Let  $IT_{PQ}$  represent the input tokens required to execute a query which are calculated as

$$IT_{PQ} = \frac{C}{A} \times n_{parse} + \frac{C}{A} \times n_{retrieve} \quad (5)$$

Where  $C$  are the characters used per query,  $A$  the average characters per token,  $n_{parse}$  the number of parsing calls from the employees and  $n_{retrieve}$  the number of retrieval calls.

Let  $T_{PS}$  denote the total number of tokens per second calculated as

$$T_{PS} = Q_{PS} \times (IT_{PQ} + OT_{PQ}) \quad (6)$$

Where  $IT_{PQ}$  are the total amount of input tokens per query and  $OT_{PQ}$  the total amount of output tokens per query used.

The number of output tokens are valued at 500 from tokenization guidelines (4 characters per token, 0.75 words per token) that were created by reputable vendors.

Let  $G_t$  represent the system's throughput demand, calculated as

$$G_t = \frac{T_{PS}}{T_{max} \times u} \quad (7)$$

Where  $T_{PS}$  is the total number of tokens per second,  $T_{max}$  is the maximum throughput and  $u$  the utilization factor.

$T_{max}$  is defined by the model of the GPU (i.e. the most used GPU for AI workloads is the NVIDIA Quadro A6000 which provides a maximum throughput of 3000 tokens per second) and the utilization factor which is estimated at 85%. The utilization factor is necessary to avoid device saturation if burst query patterns occur and possible input/output delays from networking or storage devices.

Let  $M_{Total}$  denote the total amount of memory, calculated as

$$M_{Total} = M_{Model} + \frac{m_{token} \times L_{Peak}}{1024} \quad (8)$$

Where  $M_{Model}$  is the memory required by the model,  $m_{token}$  is the average memory per token,  $L_{Peak}$  is the maximum number of tokens required to be kept in memory in gigabytes.

Let  $G_m$  represent the necessary amount of memory of the GenAI platform, calculated as

$$G_m = \frac{M_{Total}}{V_{GPU}} \quad (9)$$

Where  $V_{GPU}$  is the amount of video memory per GPU.

Finally, for calculating the hardware expenses, the platform is dimensioned based on total GPU count, which is calculated as

$$\mathbf{G} = \max(\mathbf{G}_t, \mathbf{G}_m) \quad (10)$$

Therefore, the dimensioning of the model adapts to  $\mathbf{G}$  and changes the number of hardware components of the platform, based on the rules that are illustrated in Table 1.

Table 1: Hardware Dimensioning Rules

Hardware Component	Rule
Server Chassis	If the number of GPUs supersedes 8 then another Chassis is added.
CPU	Each server chassis requires 2x28-core CPUs.
System Memory	Each additional GPU installed, four RDIMM 64 GB RAM are necessary in each server chassis.
GPU	When the number of total tokens per second exceeds 85% of the maximum throughput, another GPU is added.
Network Adapter	For every server chassis added, two network adapters are added.
OS Storage SSD	For every server chassis added, two 2 TB OS SSD storage devices are added.
Data Storage SSD	For every server chassis two 8 TB Data Storage SSDs are added.
Power Supply	For every server chassis added, two 2000W Titanium Rated power supplies are added.
Cooling	For each server chassis added one Liquid Assisted Cooling system is added.
UPS	For every two server chassis, one Uninterrupted Power Supply (UPS) is required.
InfiniBand Switch	If there is more than one server chassis, an InfiniBand Switch is required to distribute the load between all the available GPUs.
QSFP Cables	If there is more than one server chassis, 4 QSFP cables per extra server chassis are added.
Miscellaneous	Various other static costs such as patch cables, spare parts and consumables.

Let  $C_i$  represent the cost per item, calculated as

$$C_i = Q_i \times I_i \quad (11)$$

Where  $Q_i$  is the quantity of each item and  $I_i$  is the cost per item.

As a result, total hardware cost  $C_{HW}$  is calculated as

$$C_{HW} = \sum_i C_i \quad (12)$$

For the software CAPEX denoted as  $C_{SW}$ , we consider the one-time costs of licenses and optional enterprise tools for faster deployment. Therefore, it is calculated as

$$C_{SW} = \sum_j I_j \quad (13)$$

Hence, the combined hardware and software cost  $C_{HWSW}$  is calculated as

$$C_{HWSW} = C_{HW} + C_{SW} \quad (14)$$

Labor costs are the human capital costs associated with the development of the GenAI platform, the preparation of the enterprise data, the auditing process of the platform for possible risks, the project

management and the training of the employees. Table 2 presents the required roles and their descriptions in detail.

Table 2: Description of required labor roles.

Role	Description
Developer	GenAI development using frameworks, AI model integration into enterprise applications.
Data Engineer	Data pipeline design for GenAI training and data quality assurance across multiple sources.
Data Scientist	GenAI model Fine-Tuning, enterprise data migration and multimodal AI solution development.
Office Specialist	Administrative tasks of AI adoption and documentation handling
Quality Assurance and Risk Manager	AI model performance audit, bias mitigation strategy implementation, cybersecurity risks in integration and automation reviewing.
Platform Engineer	Machine Learning Operations implementation and infrastructure management.
Project Manager	Cross-functional teams management, agile framework implementation and project budgeting.

Let  $C_r$  represent the cost per role, calculated as

$$C_r = HLC_r \times H \times D \times M_r \times FTE_r \quad (15)$$

Where  $HLC_r$  is the hourly labor cost of each role,  $M_r$  is the duration required in months, and  $FTE_r$  the full-time equivalent coefficient.

For the CAPEX Build scenario, the necessary roles are Developer, Data Engineer, Data Scientist, Risk Management, Platform Engineering, and Product Management.

Let  $C_{UT}$  denote the total cost of user training which is estimated as a standard cost, calculated as

$$C_{UT} = TR_{Fee} \times U \quad (16)$$

Where  $TR_{Fee}$  is the cost of training per user.

As a result, the total labor cost ( $C_{Labor}$ ) is calculated as

$$C_{Labor} = \sum_r C_r + C_{UT} \quad (17)$$

Finally, CAPEX in the build scenario is defined as

$$CAPEX_{Build} = C_{HWSW} + C_{Labor} \quad (18)$$

The distinction of CAPEX between the buy and build scenarios is defined by not including the  $C_{HWSW}$  variable. Equation (18) defines CAPEX when building a GenAI solution; therefore, CAPEX for the buy scenario is calculated as

$$CAPEX_{Buy} = C_{Labor} \quad (19)$$

Furthermore, the evaluated use-cases require different roles, with each scenario requiring distinct role configurations based on its operational requirements. In all use-cases, all roles are necessary, except for the role of Office Specialist in the use-case of Computer Programming, the focus of GenAI platforms is



technical, and all tasks managed by an Office Specialist in other use-cases can be completed by a developer.

### 2.3 Recurring Costs of GenAI (OPEX Calculation)

Operating expenditures are the recurring expenses after the initialization of the GenAI platform. These expenses and their description are listed in Table 3, showcasing whether they apply to the build or the buy scenario.

Table 3: Operating Expenses

Operating Expense	Build	Buy	Description
User Licensing	X	✓	Licenses that authorize employees to use a GenAI platform
Token Based Cost Consumption	X	✓	The cost of all queries generated from employees
AI Tool Licensing	✓	✓	License for enterprise tools that provide automation with AI/ML capabilities, process optimization and analytics.
Power Consumption	✓	X	Electrical power consumed from a GenAI platform used
Maintenance	✓	X	Necessary maintenance during the GenAI platform’s lifecycle.
Data Management	✓	✓	Collecting, organizing, protecting and storing an enterprise’s data

Maintenance includes and supports hardware and software components, monitoring and logging services, cooling expenses, and the labor costs of system administrators, development operations engineers, and data scientists who ensure that the model remains performant, secure, and compliant over its service life. OPEX in both scenarios includes the AI tool licensing fees, the maintenance costs and the data management cost. When considering the build scenario, OPEX accounts for the costs of yearly fine-tuning the GenAI model and the cost of power consumption.

The maintenance costs are listed in Table 4, and refer to the expenses that will occur during the operation of the GenAI model.

Table 4: Maintenance Expenses

Type of Maintenance	Description
Hardware	Annual Servicing GPU Clusters, servers and cooling systems
Software Updates	LLM version upgrades and security patches
Data Pipeline	Vector database optimization and process monitoring
Compliance Audits	GDPR/HIPAA compliance checks and bias testing
Performance Optimization	Query latency monitoring and GPU utilization tuning
Security Monitoring	API Endpoints protection and Adversarial attack prevention

Let  $MC$  represent the maintenance costs, calculated as

$$MC = CAPEX \times OM \quad (20)$$

Where  $OM$  is the ongoing maintenance factor, denoting that  $MC$  is a percentage of the total CAPEX.

For the AI Tool Licensing fees, multiple tools that aid model development, model specialization for each use-case, deployment and monitoring of the platform are considered.

Let  $AT$  represent the total AI Tool Licensing fees, calculated as

$$AT = \frac{AI_{Tools \text{ Licensing per } 1000 \text{ Users}}}{1000} \times U \times M \quad (21)$$

Where  $AI_{Tools \text{ Licensing per } 1000 \text{ Users}}$  is the cost per 1,000 users, implying that the total fee scales with the total number of users that will use the specific model.

Finally, let  $DM$  represent the costs associated with data management, calculated as

$$DM = AT \times RI_{DATA \text{ MANAGEMENT}} \quad (22)$$

Where  $RI_{DATA \text{ MANAGEMENT}}$  is the ratio which new data is added to the GenAI platform and deployed properly to minimize any negative impact of previously learned knowledge.

### 2.3.1 Build Scenario

In the build scenario we consider the power consumption cost and the cost of fine tuning the model. Towards this end, let  $H_{Load}$  represent the load hours where the GenAI platform is in use by its employees and is calculated as

$$H_{Load} = H \times D \times M \quad (23)$$

Similarly, let  $H_{Idle}$  denote the total hours in all non-working days within a calendar year and the remaining non-working hours within a working day, calculated as

$$H_{Idle} = \{M \times [(365 - D) \times 24]\} + [(24 - H) \times D \times M] \quad (24)$$

Let  $PC$  represent the total amount of power consumption by the GenAI platform, calculated as

$$PC = PC_{Load} \times H_{Load} + PC_{Idle} \times H_{Idle} \quad (25)$$

Where  $PC_{Load}$  is the yearly maximum power consumption the GenAI platform will consume during working hours,  $PC_{Idle}$  is defined as the power consumption when the platform is idle for a calendar year. Both  $PC_{Load}$  and  $PC_{Idle}$  are dependent on the dimensions of the platform.

Moreover, let  $C_{PS}$  denote the total cost of power consumption, equal to the cost of energy at the location where the GenAI platform is built ( $E_{Cost}$ ). Within the techno-economic model we used the median cost of electrical energy within the European Union which is 0.22€ per kWh [14].  $C_{PS}$  is calculated as

$$C_{PS} = PC \times E_{Cost} \quad (26)$$

Fine Tuning is the process of deploying a pretrained model and training it on a smaller targeted dataset to specialize the GenAI for a particular use-case. Let  $C_{FN}$  represent the cost of the fine-tuning process, calculated as

$$C_{FN} = \sum_r C_r \quad (27)$$

Where  $C_r$  denotes the cost of the different roles, i.e. the different cost of a developer, data engineer and a data scientist.

Finally, let  $OPEX_{Build}$  represent the total operational expenses of the build scenario, calculated as

$$OPEX_{Build} = MC + AT + DM + C_{PS} + C_{FN} \quad (28)$$

### 2.3.2 Buy Scenario

Within the buy scenario, the same use-cases were evaluated as in the  $CAPEX_{Buy}$ . In addition to the default operational costs mentioned previously, the Buy scenario must include licensing fees for the employees that will use the GenAI platform, token-based consumption cost per year while removing  $MC$ .

Licensing is dependent on the number of users, the licensing cost per user and the duration the GenAI platform will operate.

Let  $L$  represent licensing, calculated as

$$L = U \times \frac{L_{per\ 1000\ Users}}{1000} \times M \quad (29)$$

Where  $L_{per\ 1000\ Users}$  is the cost per 1,000 users, implying that the total fee scales with the total number of users that will use the platform.

Token based cost consumption is calculated from the number of queries each user performs every day, the cost for each query,  $D$  and  $M$ .

Let  $CQ$  represent the cost of each query, calculated as

$$CQ = UI + UO \quad (30)$$

Where  $UI$  is the cost of a user's input and  $UO$  the cost of user's output.

Let  $TB_{CC}$  represent token-based cost consumption, calculated as

$$TB_{CC} = U \times CQ \times NC_{PD} \times D \times M \quad (31)$$

Where  $NC_{PD}$  is the number of queries each user performs daily.

Equation (32) shows how OPEX is modified to accommodate the needs of Bought GenAI.  $OPEX_{Buy}$  is calculated as

$$OPEX_{Buy} = AT + DM + TB_{CC} + L \quad (32)$$

## 2.4 Total Cost

The **Total Cost (TC)** for bought GenAI and built GenAI in an enterprise is defined from the calculated **CAPEX**, the calculated **OPEX** (either scenario) and the amount of years the infrastructure or service will be used.

Let  $TC$  represent total cost for a GenAI platform and is calculated as

$$TC = CAPEX + (OPEX * Y) \quad (33)$$

Where  $Y$  is the number of years the infrastructure or service will be used.

For a build scenario, an enterprise's infrastructure is upgraded every 3-5 years since typical infrastructure lifecycles for AI hardware are considered within a range of 3-5 years due to rapid technological advancements aligning with observations on hardware reliability trends [13]. For a buy scenario every 3 years is a recommended practice since it aligns with refresh cycles for model performance and compliance.

### 3 Results

To apply the proposed methodology, Table 5 presents the values and ranges for each parameter used as an input.

Table 5: Input Parameters

Parameters	Value	Description
$M_r$	1	The number of months necessary to create a GenAI
$M$	12	The number of months within a year.
$D$	20	The number of workdays within a month
$H$	8	The number of hours within a workday.
$T_{Out}$	500	Output tokens
$T_{max}$	3000	Maximum Throughput of a GPU
$u$	85%	GPU Utilization factor.
$P_{GAINS}$	10%-55%	Percentage of productivity gains
$P_{LEAK}$	20%-30%	Percentage of productivity leak
$OM$	10%	Ongoing Maintenance
$RI_{DATA\ MANAGEMENT}$	30%	Ratio which new data is added
$NC_{PD}$	150	Number of queries per user per day
$HLC_r$	50-100	Hourly labor cost of each role in €
$E_{Cost}$	0.22	Cost of energy per kWh in €
$U$	1-3000	Number of employees in an enterprise

The results of the model provide a 4-year study period comparison between the selected use cases for buying a GenAI platform and building a GenAI platform. In the model, the independent variable was the number of employees, with values tested from 0 to 3,000 at intervals of 20. In each case, a table with the number of employees, the costs of both the build and buy scenarios, and the value for each use case is provided, along with a line chart for up to 260 employees and a line chart for up to 3000 employees. The selection of data from 1 to 260 employees was based on the EU Enterprise Size categories, where a micro-enterprise has up to 10 employees and a large enterprise has over 250 employees [13].

#### 3.1 GenAI Computer Programming use case

Table 6 provides the costs of buying a GenAI and Building a GenAI up to 260 employees for Computer Programming.

Table 6: Build vs Buy costs and value of GenAI for Computer Programming case (1 – 260 Employees)

Employees	Buy (€)	Build (€)	Value (€)
1	25,196	186,980	31,094
20	318,063	213,865	621,881
40	631,707	264,007	1,243,761
60	940,922	297,015	1,865,642
80	1,245,710	312,888	2,487,523
100	1,550,497	328,760	3,109,404
120	1,855,285	344,633	3,731,284
140	2,160,072	372,572	4,353,165
160	2,464,860	388,445	4,975,046
180	2,769,647	404,318	5,596,927
200	3,074,434	420,190	6,218,807
220	3,379,222	436,063	6,840,688
240	3,684,009	451,936	7,462,569
260	3,988,550	466,093	8,084,450

Comparing the increase in costs between building and buying a GenAI platform for up to 3,000 employees in Computer Programming as shown in Figure 3, we can see how the expenses for both scenarios show a linear increase while the Build scenario increases at a much lower rate.

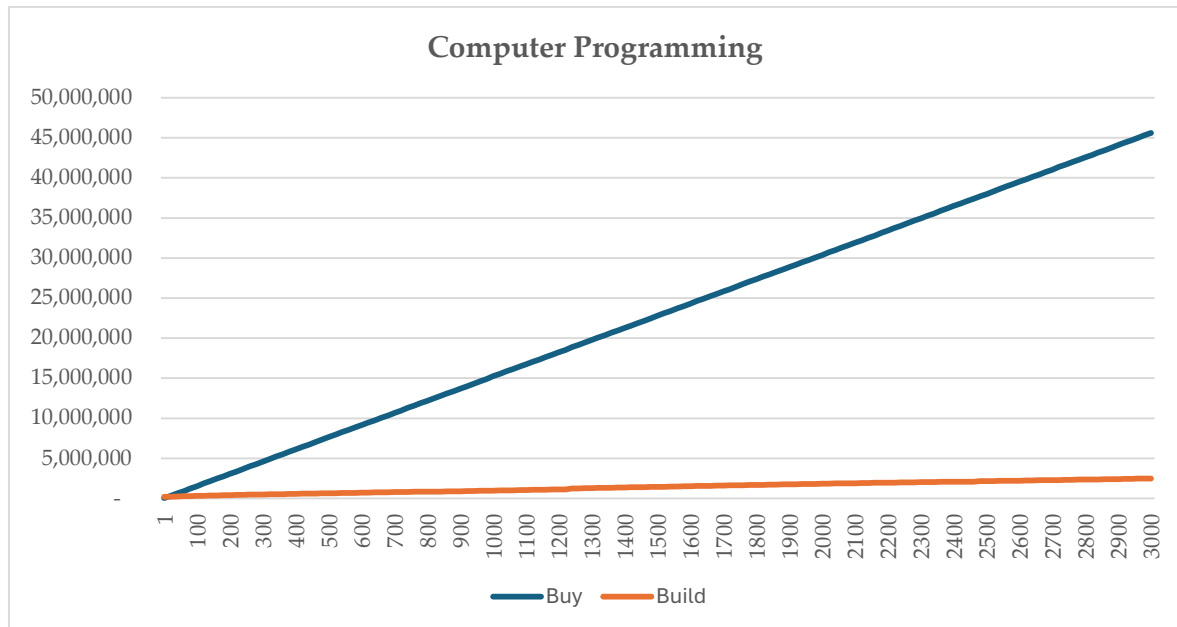


Figure 3: Build vs Buy costs up to 3000 employees in Computer Programming

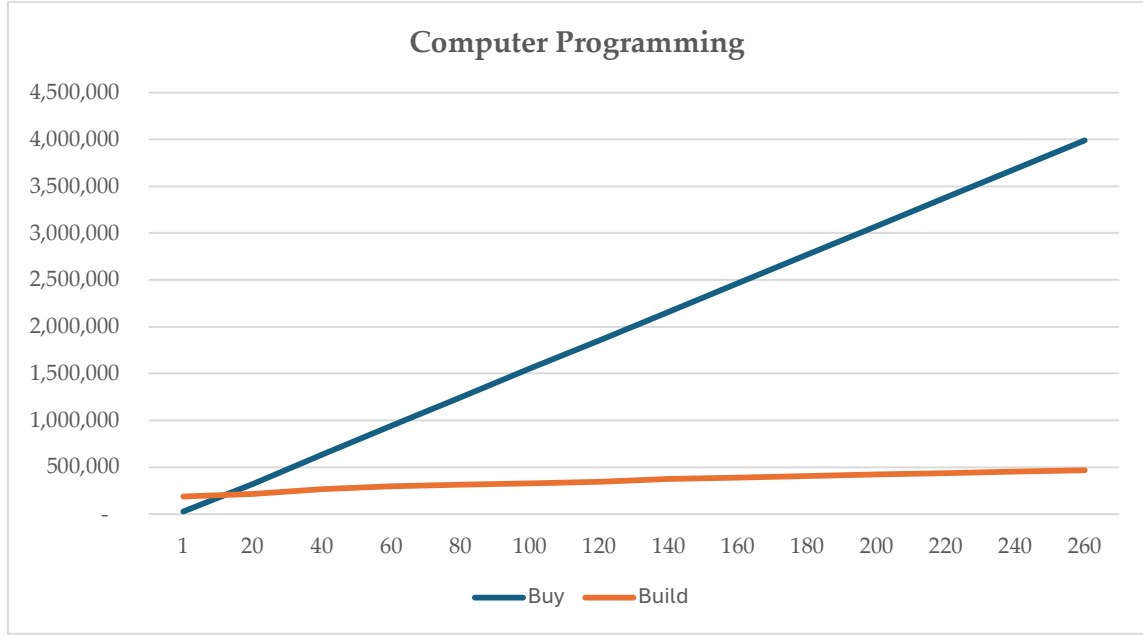


Figure 4: Build vs Buy costs up to 260 employees in Computer Programming

In Figure 4 we notice how costs increase depending on the number of employees. We also notice that costs for Building a GenAI platform for Computer Programming is cost-effective compared to Buying at 13 employees.

### 3.2 GenAI Business Productivity use case

Table 7 provides the costs of Buying a GenAI scenario and Building a GenAI scenario up to 260 employees for Business productivity.

Table 7: Build vs Buy from 1 – 260 Employees in Business Productivity

Employees	Buy (€)	Build (€)	Value (€)
1	12,126	191,800	3,317
20	55,091	219,365	66,334
40	104,202	270,867	132,668
60	150,875	305,927	199,002
80	195,110	324,545	265,336
100	239,346	343,163	331,670
120	283,581	361,781	398,004
140	327,816	392,465	464,338
160	372,051	411,082	530,672
180	416,286	429,700	597,006
200	460,521	448,318	663,339
220	504,756	466,936	729,673
240	548,991	485,553	796,007
260	592,894	501,517	862,341

According to Figure 5, both scenarios implement a linear increase while the Build scenario increases at a lower rate compared to the Buy scenario.

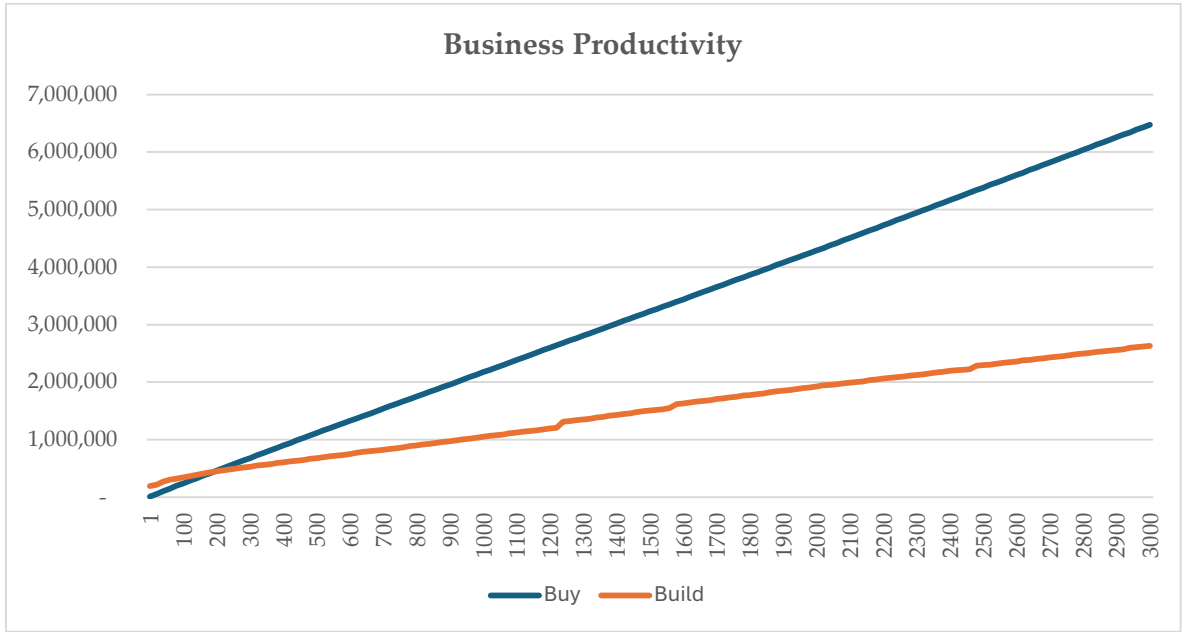


Figure 5: Build vs Buy costs up to 3000 employees in Business Productivity.

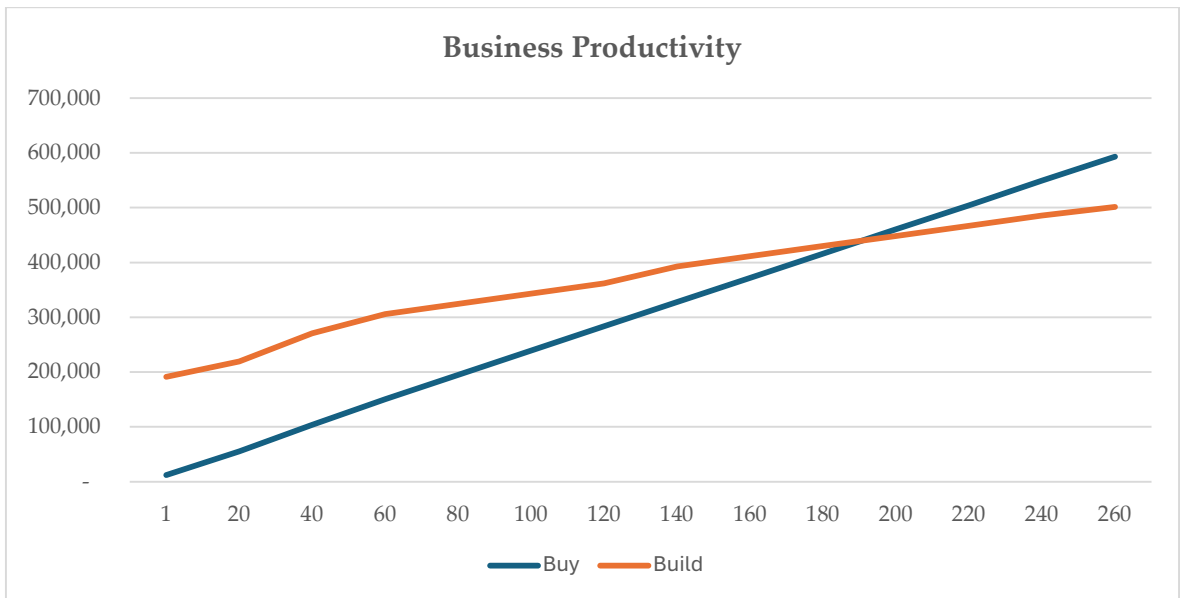


Figure 6: Build vs Buy costs up to 260 employees in Business productivity

A Built GenAI platform for Business Productivity tasks becomes less costly compared to the Buy scenario at 195 employees, as shown in Figure 6.

### 3.3 GenAI Customer Support use case

Table 8 provides the costs of Buying a GenAI scenario and Building a GenAI scenario up to 260 employees for Customer Support.

Table 8: Build vs Buy from 1 – 260 Employees in Customer Support

Employees	Buy (€)	Build (€)	Value (€)
1	13,556	191,000	4,018
20	45,577	218,055	80,366
40	84,494	268,536	160,732
60	121,475	301,586	241,098
80	156,520	317,205	321,465
100	191,565	332,824	401,831
120	226,610	348,443	482,197
140	261,655	376,128	562,563
160	296,700	391,747	642,929
180	331,744	407,366	723,295
200	366,789	422,985	803,661
220	401,834	438,604	884,027
240	436,879	454,223	964,394
260	471,211	468,683	1,044,760

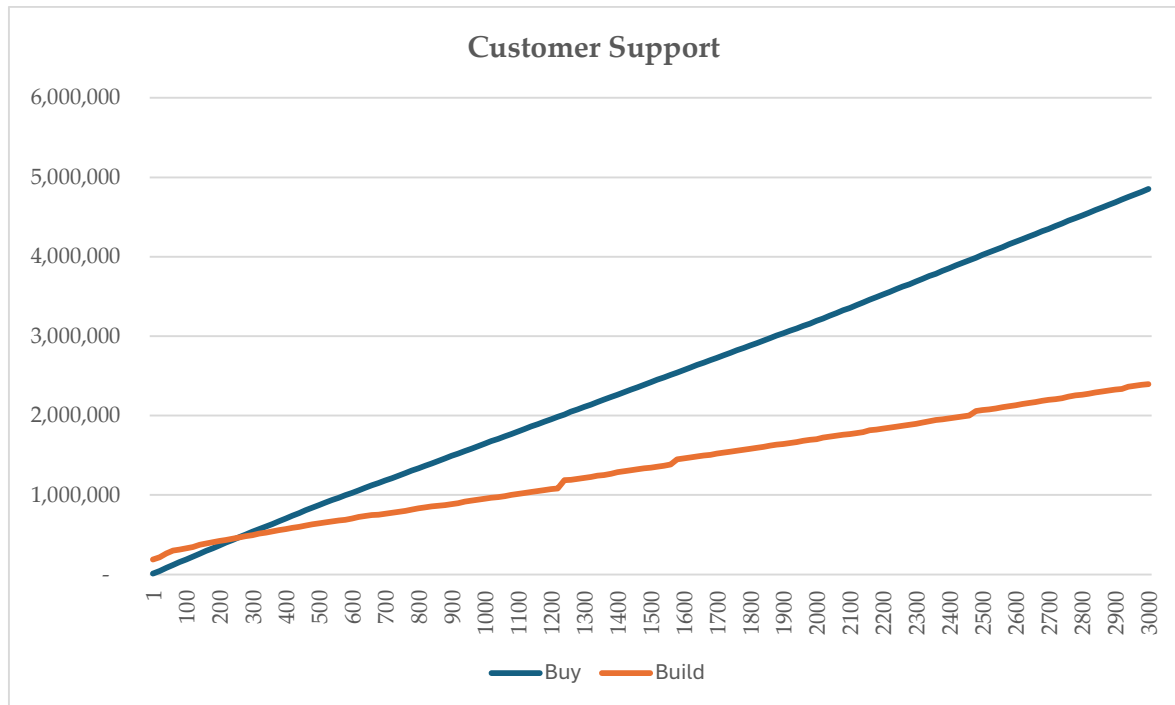


Figure 7: Build vs Buy costs up to 3000 employees in Customer Support

As has been shown in all previous use-case scenarios, Figure 7 shows that both scenarios have a linear increase as the number of employees increase while the Build scenario maintains a lower rate.



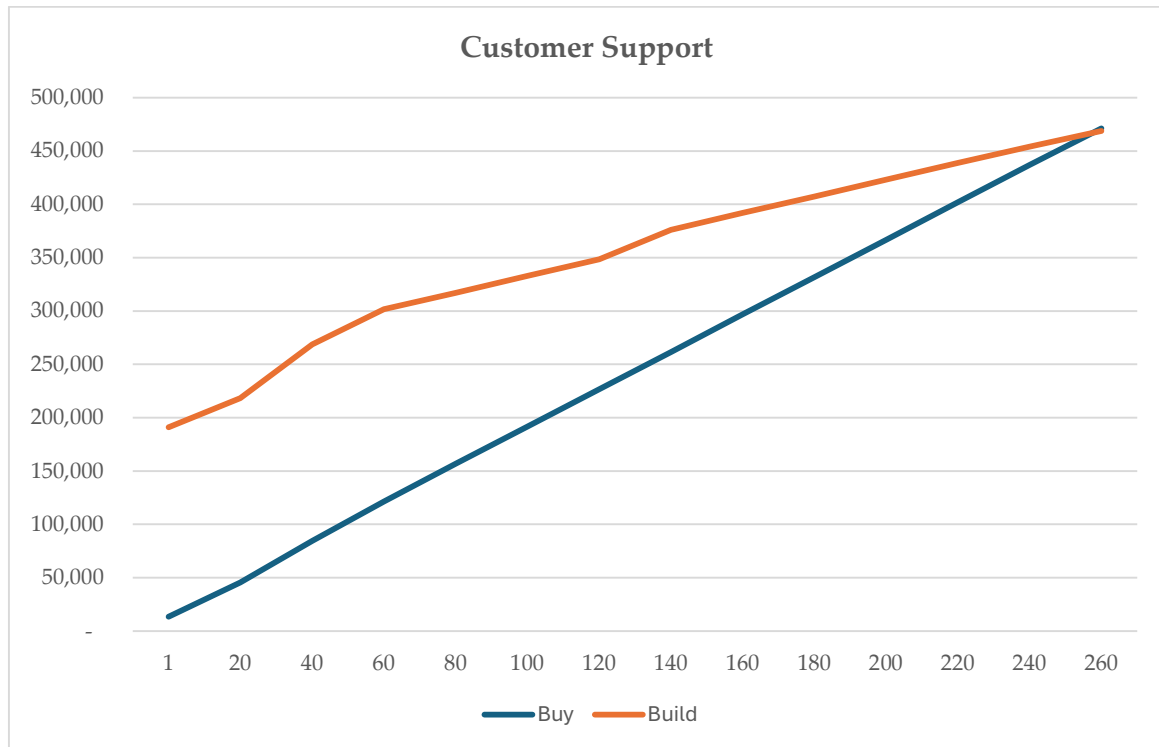


Figure 8: Build vs Buy costs up to 260 employees in Customer Support

The cost of a Build scenario for Customer Support is less than the Buy scenario for 258 employees and above as shown in Figure 8.

### 3.4 GenAI Marketing use case

Table 9 provides the costs of Buying a GenAI scenario and Building a GenAI scenario up to 260 employees for Marketing.

Table 9: Build vs Buy from 1 – 260 Employees in Marketing

Employees	Buy (€)	Build (€)	Value (€)
1	24,156	191,000	31,253
20	319,343	218,055	625,070
40	637,627	268,536	1,250,140
60	950,776	304,077	1,875,210
80	1,258,792	324,679	2,500,279
100	1,566,807	345,281	3,125,349
120	1,874,823	365,883	3,750,419
140	2,182,838	398,552	4,375,489
160	2,490,854	419,154	5,000,559
180	2,798,869	439,755	5,625,629
200	3,106,884	460,357	6,250,699
220	3,414,900	480,959	6,875,769
240	3,722,915	501,561	7,500,838
260	4,030,362	519,178	8,125,908

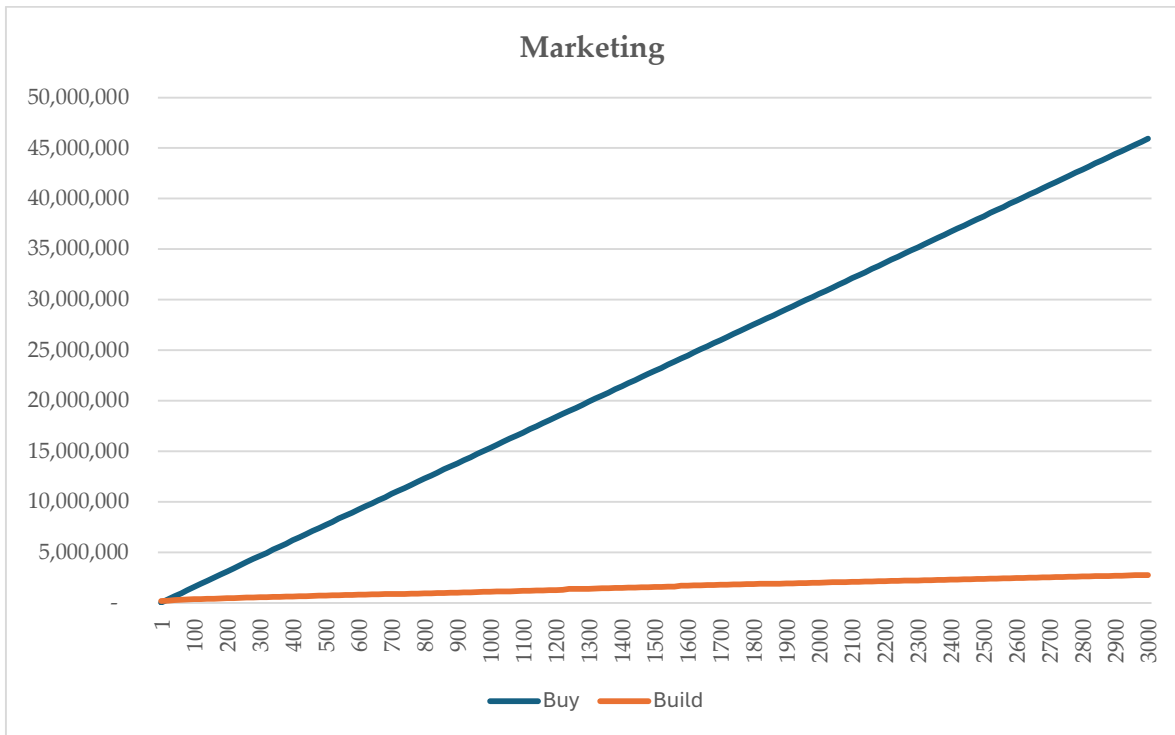


Figure 9: Build vs Buy costs up to 3000 employees in Marketing.

According to Figure 10 the Build scenario costs less, at 13 employees compared to the Buy scenario and finally the Build scenario remains less costly up to 3000 employees as shown in Figure 9.

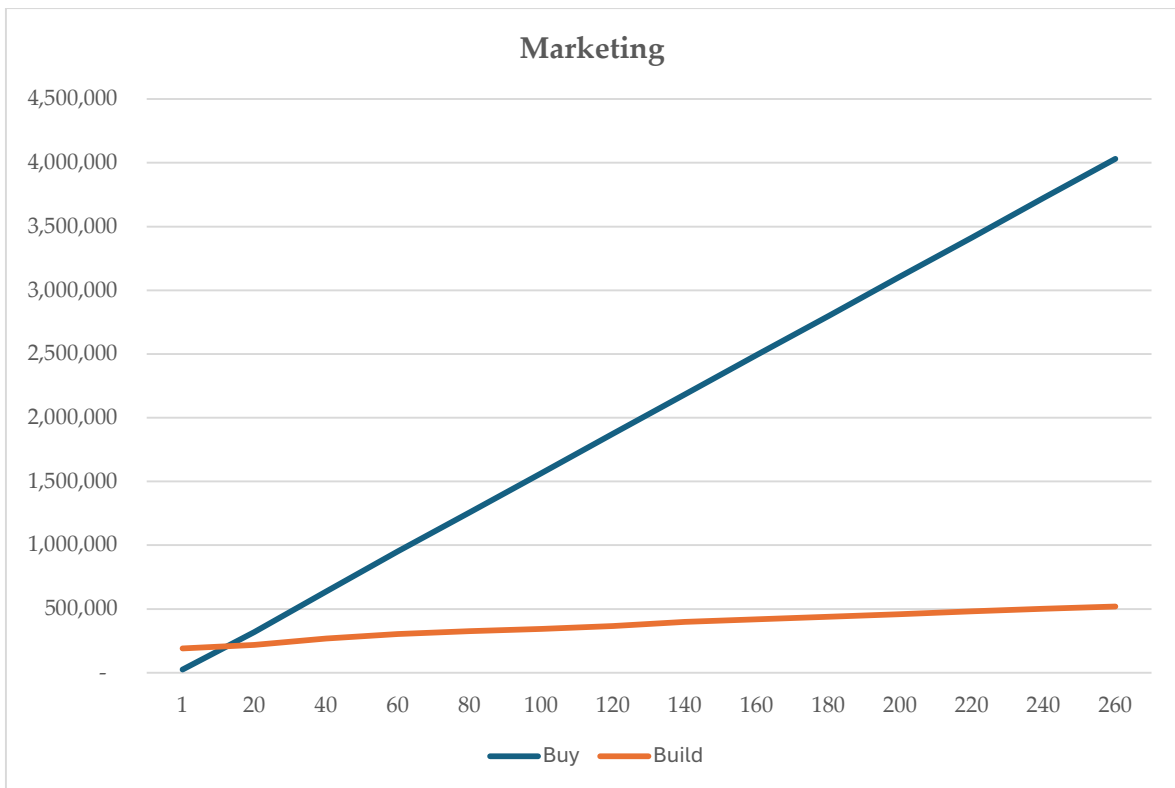


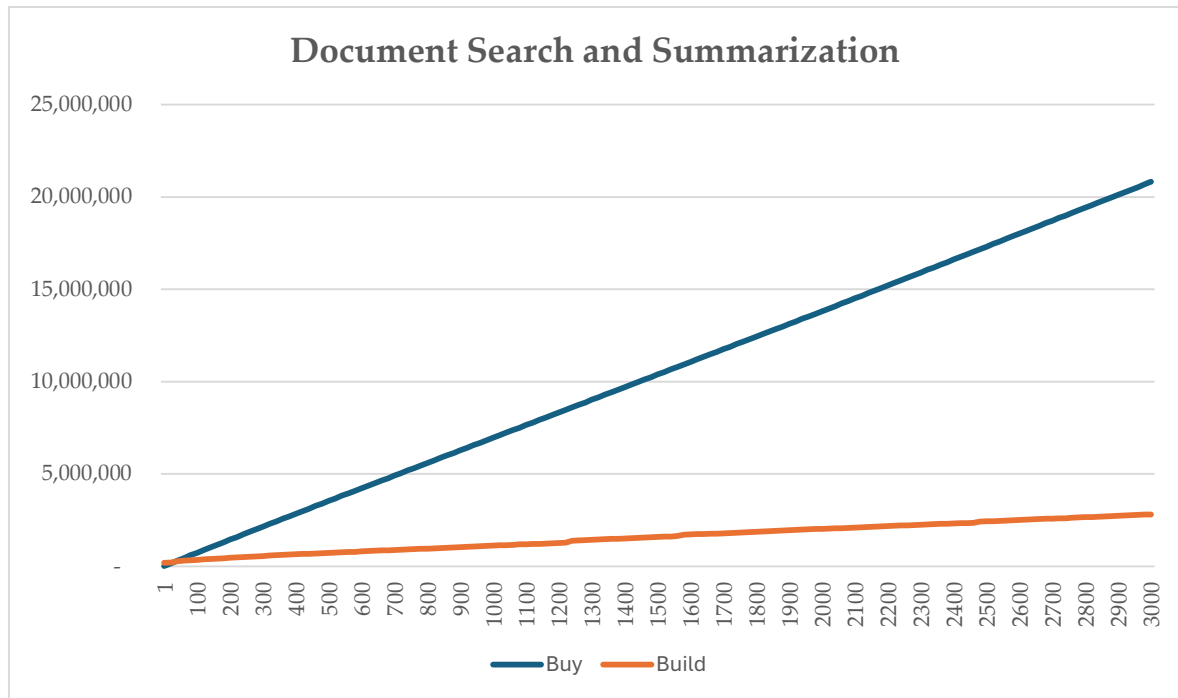
Figure 10: Build vs Buy costs up to 260 employees in Marketing

### 3.5 GenAI Document Search and Summarization use case

Table 11 provides the costs of Buying a GenAI scenario and Building a GenAI scenario up to 260 employees for Document Search and Summarization.

*Table 10: Build vs Buy from 1 – 260 Employees in Marketing*

Employees	Buy (€)	Build (€)	Value (€)
1	21,908	192,120	12,757
20	158,223	220,344	255,131
40	310,107	273,164	510,261
60	456,756	309,963	765,392
80	598,172	330,741	1,020,522
100	739,587	351,519	1,275,653
120	881,003	372,297	1,530,783
140	1,022,418	405,141	1,785,914
160	1,163,834	425,918	2,041,044
180	1,305,249	446,696	2,296,175
200	1,446,665	467,474	2,551,306
220	1,588,080	488,252	2,806,436
240	1,729,496	509,029	3,061,567
260	1,870,912	527,202	3,316,697



*Figure 11: Build vs Buy costs up to 3000 employees in Document Search and Summarization.*

Using a GenAI platform for Document Search and Summarization, a build scenario becomes cost efficient at 33 employees as shown in Figure 12, while at Figure 11 we notice a linear increase with a higher rate for the Buy scenario.

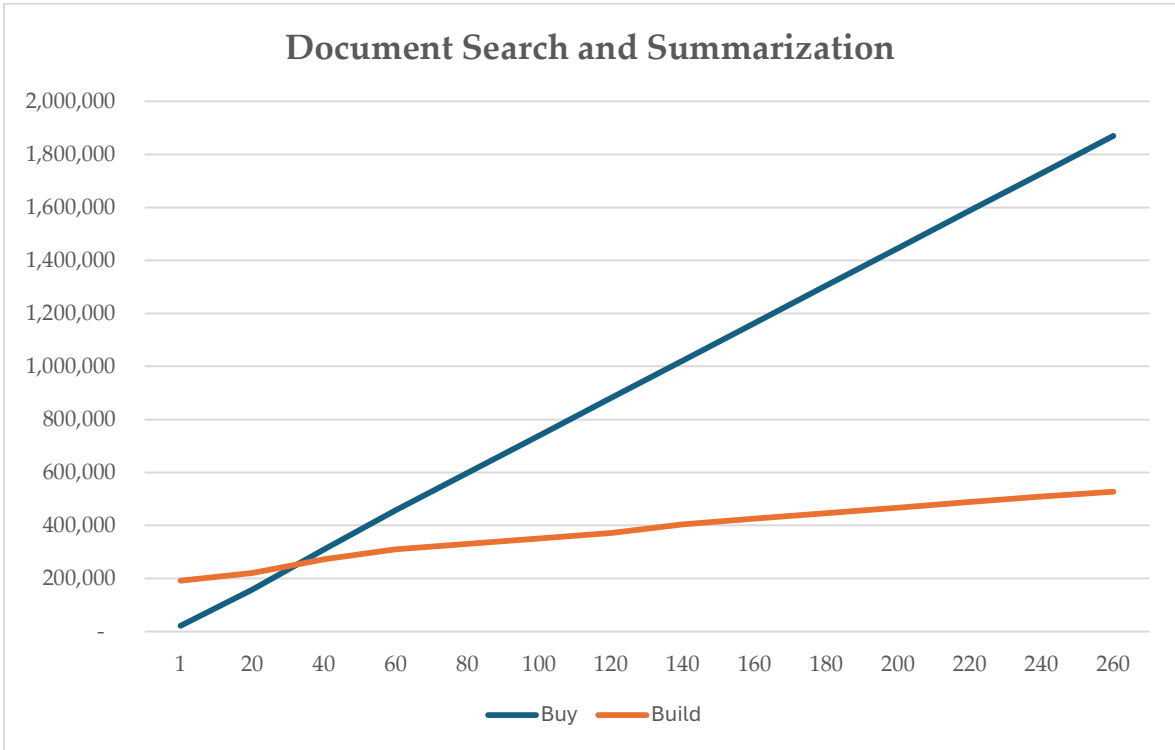


Figure 12: Build vs Buy costs up to 260 employees in Document Search and Summarization

#### 4 Discussion & Conclusion

In this study, we developed a decision-making model to guide enterprises in evaluating the economic trade-offs between building a custom GenAI solution and purchasing a GenAI service. Our results demonstrate that the viability of building a proprietary GenAI platform is highly dependent on the number of employees. As the duration of infrastructure use increases, the threshold number of employees required to justify an in-house solution decrease, primarily due to the amortization of initial capital expenditures over a longer period.

In all cases, micro enterprises appear to have no benefit in building a GenAI platform, while small, medium and large enterprises may benefit in Computer Programming, Marketing and Document Search and Summarization use cases. In the Customer Support use case, only large enterprises may benefit from building a GenAI and in Business Productivity medium and large enterprises may reduce their costs through a Build scenario as shown in Table 11 .

Table 11: Build or Buy selection depending on the size of an Enterprise

Enterprise size	Number of Employees	Computer Programming	Business Productivity	Customer Support	Marketing	Document Search and Summarization
<b>Micro</b>	0 - 9	Buy	Buy	Buy	Buy	Buy
<b>Small</b>	10 -49	Buy	Buy	Buy	Build	Build
<b>Medium</b>	50 - 249	Build	Build	Buy	Build	Build
<b>Large</b>	>249	Build	Build	Build	Build	Build

The model quantifies value primarily through labor-time savings. While other strategic benefits (e.g., enhanced innovation or competitive advantage) or risks (deployment/implementation challenges or vendor lock-in) while they do exist, they are not part of the quantitative calculations.

The number of active work hours per day and active workdays per year may not be accurate in all dynamic business environments.

Multiple factors must be considered other than economic thresholds and productivity gains. Local laws and regulations must be adhered to. General Data Protection Regulation (GDPR) for EU enterprises that handle data of EU citizens, Health Insurance Portability and Accountability Act (HIPAA) for US enterprises that manage medical data, or any enterprise that owns intellectual property, have a requirement to keep their data on-premises or under a dedicated cloud account instead of sending data to a third-party GenAI vendor [16]. For this reason, fine-tuning was studied only for build scenarios instead of buy.

These results offer practical insights for enterprises seeking to optimize their AI-adoption strategies. Enterprises with many employees and long-term AI integration plans may benefit from investing in custom-built GenAI, whereas smaller enterprises or those with shorter planning horizons may find greater value in purchasing GenAI services from a reputable vendor. Beyond cost considerations, decision-makers should also factor in qualitative parameters such as data sovereignty, regulatory compliance, and the availability of required personnel to implement the AI adoption strategy.

Finally, the model may become a practical tool for enterprises to make informed, financially based decisions regarding GenAI integration. Future research may extend this model to incorporate more sectors within an enterprise, evolving AI capabilities and technologies, and dynamic pricing structures, further enhancing its applicability across diverse organizational contexts. Finally, the adjustment of the model in a per country basis using the available productivity data, labor costs and mix of types of enterprises could allow for estimating adoption of GenAI and identify barriers or incentives in each country.

## References

- [1] S. Noy and W. Zhang, 'Experimental evidence on the productivity effects of generative artificial intelligence', *Science*, vol. 381, no. 6654, pp. 187–192, 2023.
- [2] Ni, Xiao and Wang, Yiwei and Feng, Tianjun and Lu, Lauren Xiaoyuan and Wang, Yitong and Zhou, Congyi, Generative AI in Action: Field Experimental Evidence on Worker Performance in E-Commerce Customer Service Operations (November 06, 2024). Available at SSRN: <https://ssrn.com/abstract=5012601>
- [3] T. Besiroglu, N. Emery-Xu, and N. Thompson, 'Economic impacts of AI-augmented R&D', *Research Policy*, vol. 53, no. 7, p. 105037, 2024.
- [4] Hoffreumon, C., Forman, C., & Zeebroeck, N. v. (2024). Make or buy your artificial intelligence? Complementarities in technology sourcing. *Journal of Economics & Management Strategy*, 33, 452–479. <https://doi.org/10.1111/jems.12586>
- [5] Svanberg, Maja and Li, Wensu and Fleming, Martin and Goehring, Brian and Thompson, Neil, Beyond AI Exposure: Which Tasks are Cost-Effective to Automate with Computer Vision? (January 19, 2024). Available at SSRN: <https://ssrn.com/abstract=4700751>
- [6] Acemoglu, Daron, The Simple Macroeconomics of AI (May 2024). NBER Working Paper No. w32487, Available at SSRN: <https://ssrn.com/abstract=4843046>
- [7] Fangchen Song, Ashish Agarwal and Wen Wen, The Impact of Generative AI on Collaborative Open-Source Software Development: Evidence from GitHub Copilot (October 2024). <https://doi.org/10.48550/arXiv.2410.02091>
- [8] Al Naqbi H, Bahroun Z, Ahmed V. Enhancing Work Productivity through Generative Artificial Intelligence: A Comprehensive Literature Review. *Sustainability*. 2024; 16(3):1166. <https://doi.org/10.3390/su16031166>
- [9] S. Pansombut and K. Kirimasthong, Customer Service Transformation for Admission in Mae Fah Luang University Using Chatbots Technology, 2025 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT & NCON), Nan, Thailand, 2025, pp. 378-383, doi: 10.1109/ECTIDAMTNCN64748.2025.10962037. <https://ieeexplore.ieee.org/abstract/document/10962037>
- [10] Marko Vidrih and Shiva Mayahi, Generative AI-Driven Storytelling: A New Era for Marketing (September 2023). <https://doi.org/10.48550/arXiv.2309.09048>
- [11] Nick Whitehouse, Nicole Lincoln, Stephanie Yiu, Lizzie Catterson and Rivindu Perera, Better Bill GPT: Comparing Large Language Models against Legal Invoice Reviewers, (April 2025), <https://doi.org/10.48550/arXiv.2504.02881>
- [12] Rajan Gupta, Gaurav Pandey, and Saibal Kumar Pal. Automating Government Report Generation: A Generative AI Approach for Efficient Data Extraction, Analysis, and Visualization. *Digit. Gov.: Res. Pract.* 6, 1, Article 4 (March 2025), <https://doi.org/10.1145/3691352>.
- [13] K. V. Vishwanath, N. Nagappan, Characterizing cloud computing hardware reliability (June 2010). Proceedings of the 1st ACM symposium on Cloud computing. <https://doi.org/10.1145/1807128.1807161>
- [14] Eurostat, Electricity prices for household consumers - bi-annual data (from 2007 onwards), [https://ec.europa.eu/eurostat/databrowser/view/nrg\\_pc\\_204/default/table?lang=en](https://ec.europa.eu/eurostat/databrowser/view/nrg_pc_204/default/table?lang=en)
- [15] Glossary: Enterprise size, [https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Enterprise\\_size](https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Enterprise_size)
- [16] Gräf, Miriam; Mehler, Maren; and Ellenrieder, Sara, "AI Strategy in Action: A Case Study on Make-or-Buy for AI-based Services" (2024). PACIS 2024 Proceedings. 8. [https://aisel.aisnet.org/pacis2024/track01\\_aibussoc/track01\\_aibussoc/8](https://aisel.aisnet.org/pacis2024/track01_aibussoc/track01_aibussoc/8)