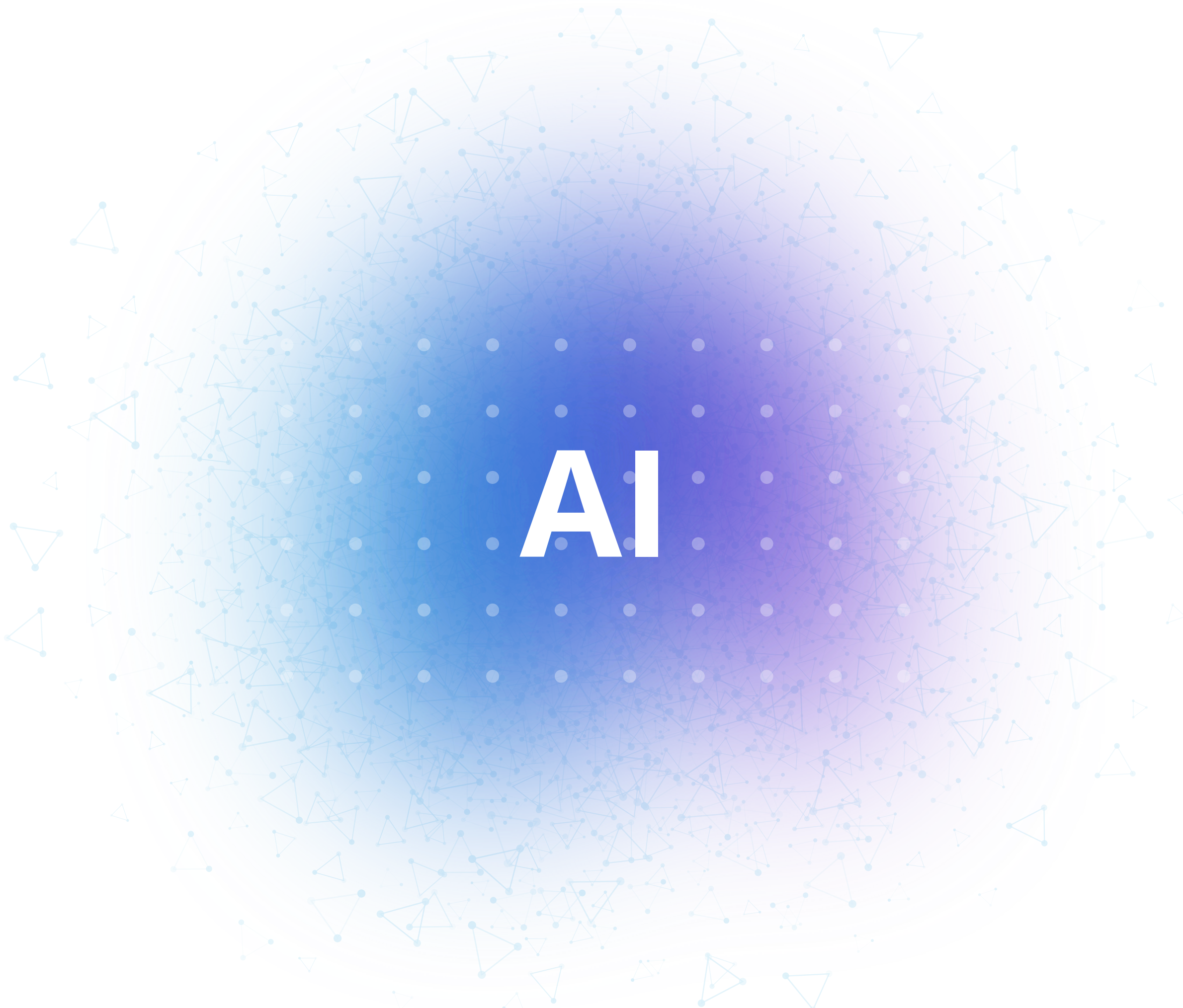
FRUGAL AI HUB at UNIVERSITY OF CAMBRIDGE Judge Business School

UNICC | Artificial Intelligence



From Total Cost of Ownership to Social Impact:
A Frugal AI Framework to Measure Your AI Portfolio as a Strategic Asset

December 2025

1. Acknowledgements

Ian Colliar, Maria Jover, Mikael Liljebladh, Shambhavi Mohan, Yaman Saqqa, Tristan Gitman,
Anhad Ahluwalia, Emile Dos Santos Ferreira, Elizabeth Osta, Serish Venkata Gandikota, Meghana Prakash, J

2. Abstract

The comprehensive measurement of the value of an Artificial Intelligence (AI) portfolio remains a significant challenge within organisations. Assessments are often nascent and fragmented, failing to account for the full spectrum of costs and impacts. This paper proposes a standardised, three-level framework for measuring AI value, developed in partnership between the Frugal AI Hub at the University of Cambridge and the United Nations International Computing Centre (UNICC).

The framework's metrics are designed for both the individual AI model and the overall portfolio in the context of an organisation or entity 'users' of AI . Level 1 focusses on the Total Cost of Ownership (TCO) and the Frugal AI metrics to quantify financial and energy costs. Level 2 introduces a return-on-investment (ROI) measurement that links financial benefits to TCO. Finally, Level 3 aligns AI performance with the Sustainable Development Goals (SDGs) of the United Nations to measure social impact. By integrating these metrics, the framework enables organisations to move beyond siloed assessments, optimising their AI portfolios for efficiency and responsible deployment.

Although developed in conjunction with the UNICC, this framework is broadly applicable to any organisation seeking a transparent and equitable valuation of AI. The framework draws on a UNICC survey of two multi-agency AI use cases covering the full design-to-run lifecycle costs, which inform the cost component breakdowns and validation of regression models.

This paper is the first of a wider research program. Other papers in progress or in the pipeline focus on further refinements of the metrics, examples of operationalisation and portfolio optimisation.

These topics will be addressed at high level only as they are not the main focus of this paper.

The AI Value Chain

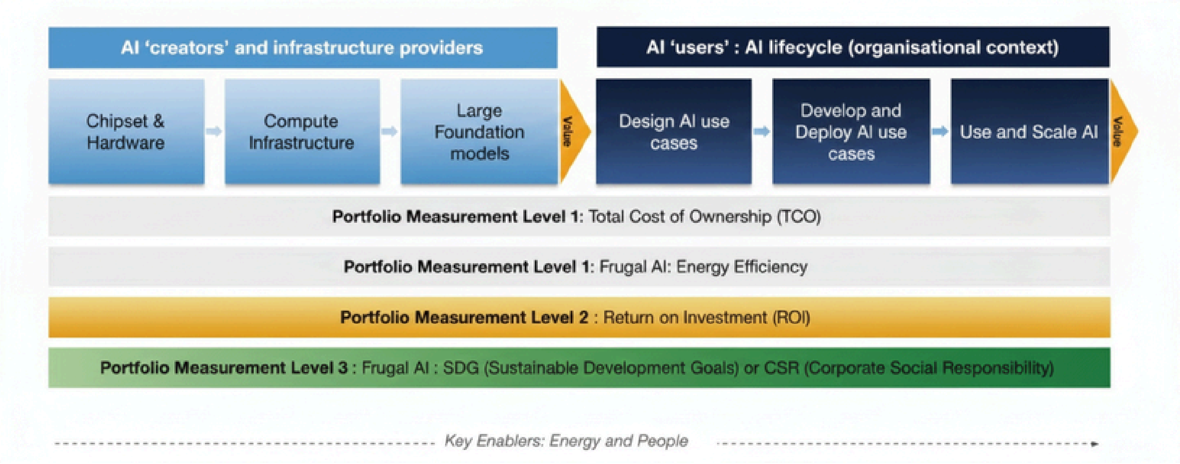


Figure 1: Illustration of the three levels of AI metrics proposed in this framework linking Total Cost of Ownership, ROI, and SDG-aligned impact for comprehensive AI portfolio evaluation.

CONTENTS

3. The Strategic Imperative for AI portfolio measurement

With a growing number of Artificial Intelligence (AI) models in production and plans to significantly expand the portfolio in organisations and entities of all sizes across the world, a robust set of metrics to measure AI portfolios along the value chain creates the transparency necessary to continuously optimise, thus achieving "more with less", a key tenet of Frugal AI.

The collaboration with United Nations International Computing Centre (UNICC) gave the opportunity to study in detail AI costs through a detailed survey. The results led to an initial approach to create transparency along the AI development cycle at model level, a fair charge back and levers to continuously optimise the AI portfolio. New AI advancements (LLM, agentic AI) unlock new opportunities as well as new cost structures, hence measurement and transparency becomes key to evaluate portfolio optimisation opportunities. This approach can be relevant to a multitude of organisations who could afford a larger number of use cases developed and scaled by optimising the cost and value of current portfolios. Many companies have adopted measures of value with different approaches. This paper focusses on a financially orientated value score for ROI measurement and introduces a set of frugal AI metrics aligned with the SDG goals to measure impact on society.

3.1 Level 1(A): A Unified TCO and Frugal AI Measurement

Arga et al. (2025) [2] examine frugal AI as a developing paradigm focused on enhancing AI systems for cost-effectiveness, environmental sustainability, and resource conservation. The framework emphasises strategies including model compression, energy-efficient architectures, and reduced data and computational needs, thereby aligning AI design with principles of affordability and scalability. The focus on minimising costs related to development, infrastructure, inference, maintenance, deployment, and support creates a significant implicit connection to Total Cost of Ownership (TCO). By promoting more efficient models and reduced resource use, the paper clearly highlights approaches that can lower TCO curves throughout the AI lifecycle, especially in terms of operational expenses.

Total Cost of Ownership (TCO) refers to a comprehensive financial metric that encapsulates the entirety of the cost, including acquisition, development, operational and maintenance. Mathematically speaking,

$$TCO = I + O + M \tag{3.1}$$

where *I* = initial cost, *O* = operational cost, *M* = maintenance cost.

However, since operational and maintenance costs are incurred throughout the lifecycle, in practice, we can extend Equation 3.1 to model TCO as a function of time, *T*, as follows:

$$TCO_{\text{total}} = I + \sum_{t=1}^{T}(O_t + M_t) \tag{3.2}$$

where *I* represents the one-time initial investment, terms O_t and M_t represent the recurring operational and maintenance costs for each time period *t*.

This distinction is important because, as will be discussed in Section 1.3.1, the allocation of initial design and development costs (*I*) requires a different treatment (e.g., tiered allocation or amortisation), whereas operational and maintenance expenses are more predictable and can be measured period by period.

For this paper, the proposed TCO framework deconstructs the identified AI costs across the entire lifecycle. This approach ensures that all direct and indirect costs are captured, establishing a clear financial commitment required for each AI initiative or model.

3.2 Level 1(B): A Deep dive on TCO components

One of the first extensive academic frameworks for implementing Total Cost of Ownership (TCO) in purchasing decisions was proposed by Ellram (1995) [6]. Therein, TCO is defined as comprehensive costs throughout the lifecycle related to the acquisition, ownership, use, and disposal of goods or services. The definition of TCO encompasses both indirect and concealed expenses, including supplier qualification, ordering processes, inventory holding costs, warranty expenses, disposal fees, and opportunity costs. This perspective shifts purchasing practices from a narrow focus on price comparisons to a broader, analytical approach. An eight-stage framework was proposed that directs the implementation process, from defining the scope to identifying cost components to measuring costs and applying the results in decision making. This framework is derived from seven comprehensive case studies of companies, showcasing both methodological precision and practical relevance. This was pivotal in transitioning TCO from a theoretical framework to a systematic practice. The focus on lifecycle cost components, the categorisation of TCO models and the establishment of an implementation framework provided a foundation for later research, which has further developed and enhanced the concept in various industries and methodological perspectives.

A critical view of the cost components is a crucial aspect of calculating the TCO. After reviewing published papers in academia and in the commercial world (Gartner, Forrester, Microsoft and others) the authors developed a detailed survey to understand cost structure and value measurement. The survey is structured into four parts, beginning with a **high-level overview** of costs, budgets, and existing metrics across the three phases of the AI lifecycle: Design, Pilot, and Scale. This first section also delves into current and future **charge-back** mechanisms for AI costs, the categorization of **fixed vs. variable costs** in model design, and cost estimates for the entire portfolio and a specific Example Use Cases. The second part focuses on collecting detailed, aggregate, and use-case-specific cost estimates for each AI lifecycle phase, probing for **critical cost levers, metrics used,** and **charged-back costs.** Part 3 shifts to the **landscape of AI models**, examining future footprint estimates, the scope of a **Total Cost of Ownership (TCO) Calculator**, and infrastructure-related levers to reduce TCO, such as cloud strategy, compute resources, and data storage. The final section is an initial discovery phase focusing on the goals and potential barriers of a new AI charging model, definition of **Value and UN Sustainability Indicators** aligned to AI metrics. The survey was compiled by UNICC for two Example Use Cases, at different stages of development. The results were analysed and aggregated into meaningful clusters, forming the basis for the proposal of key pillars to measure **Total Cost Of Ownership (TCO)** in the context of an organisation 'user' of AI.

3.2.1 Compute and Infrastructure

Cloud Services: Costs associated with cloud infrastructure, including model licences, specific DBs, document indexing, and other required components. Elasticity benchmarks are very useful (how quickly resources can be scale down when demand drops).

On-Premises: Key KPIs for consideration are benchmarks such as percentage utilisation rate per compute node, price per inference, or average idle time to model underutilised CapEx/OpEx costs. For proprietary data centres, costs related to usage of hardware and chipsets should be included here(acquisition, amortization, run costs).

Networking: Costs for data transfer, VPNs, and private links required for secure inter-agency access.

3.2.2 Data Lifecycle

Data Ingestion and Indexing: Costs associated with processing and indexing new documents for Retrieval-Augmented Generation (RAG) models like Joint AI Solution **1 and 2** and unplanned costs spikes occurring from vector search tuning, periodic re-indexing, and semantic drift tracking.

Data Preparation and Governance: Personnel and tool costs for data gathering for model training, quality, security, compliance and deletion across multiple agencies, are a critical and often underestimated expense.

Data Storage: Costs for storing raw data, processed data, and model artifacts (e.g., vector indexes for RAG).
Tracking costs for storing duplicated data allows to optimise the use of cloud estates. Moreover, practices of clear retention policies (data deletion) also contribute to cost containment.

Data Maintenance: Quantify the cost-effectiveness of data upkeep by linking the proportion of data reprocessed, its processing cost, and its measurable contribution to AI model value.

Survey Insight: Respondents highlighted that coordination and documentation costs within the data lifecycle were higher than expected, particularly during indexing and governance. These costs often exceeded initial estimates due to the complexity of multi-agency processes (see Appendix 10).

3.2.3 Models and Software

SaaS Model Usage: Recurring fees for using LLM models (inference costs and others). Example metrics: number of queries or tokens processed, or licences based on number of users.

Open-Source Alternatives: Future costs associated with deploying and maintaining opensource models and databases.

Platform and Tool Licensing: Costs for supporting software, including project management tools, monitoring dashboards, and security solutions. Here costs for other software used that it is not AI specific, for example front ends, application servers, backend logic and databases, should be included.

3.2.4 Personnel and Expertise

Development and Design: All people-costs for solution architecture, gathering of business requirements, UI/UX design, and application development. This is a primary cost driver in the Design and Pilot phases.

Stakeholder Management: Time and resources spent on inter-agency coordination, workshops, change management, and feedback loops, especially critical for multi-tenant applications like **Joint AI Solution 1 and 2.**

Support and Training: Costs for SLA-based support, hypercare, documentation, and training for partner agencies.

Survey Insight: Respondents emphasised that senior architects and subject-matter experts were the most significant personnel costs. Cross-team approvals and workshops added further overhead, reinforcing that people costs are a dominant driver in early lifecycle phases.

3.2.5 Integration, Orchestration and Decommission

System Integration: Engineering effort to connect AI services with existing agency systems and ensure secure data flow and access management.

Coordination: Overhead costs associated with developing and maintaining shared components, such as the cost-sharing model itself, which can be reused for future projects.

Decommission: A cost category to capture savings or transitional expenses from retiring legacy technologies or models replaced by new AI solutions, ensuring total lifecycle costs reflect both deployment and displacement impacts.

Survey Insight: Respondents noted that architectural complexity, particularly multi-tenant and multilingual system requirements, substantially increased integration and orchestration costs across projects.

3.2.6 Maintenance and Monitoring

Continuous Improvement (CI/CD): Costs for ongoing enhancements, development, and testing for new versions of AI services (e.g., Joint AI Solution 1 and 2).

Model and Data Upkeep: Costs incurred for updating prompts, embedding models for RAG, and re-indexing documents to prevent performance degradation (model drift).

Monitoring and Alerting: Costs of tools and personnel required to monitor system health, usage, and performance, and to generate reports and dashboards.

Survey Insight: Respondents observed that ongoing run costs were initially lower than design and development costs, but scaled steadily with increased usage, monitoring, and updates. This pattern underscores the need to model not just build-time but also growth-phase cost trajectories.

3.2.7 Governance, Risk, Compliance and Ethics

Security and Compliance: Direct costs for security reviews and compliance assessments, which are performed as one-time fixed costs during development. [13]
A timely engagement at the start of the project of Security, Compliance and Governance leads to enhanced cost effectiveness.

Data Privacy and Sovereignty: Costs to ensure adherence to inter-units data sharing agreements (SDA) and other regulatory requirements.

Ethical costs: As highlighted in a later chapter, there may be upfront costs, for example for adapting models to ensure accessibility.

Risk Management: Implementing frameworks such as the NIST AI Risk Management Framework (AI RMF) to ensure responsible and trustworthy AI deployment [13].

Survey Insight: Respondents associated governance costs not only with formal compliance reviews, but also with approval bottlenecks and risk assessment cycles across agencies. These processes, while necessary for accountability, were frequently cited as adding significant overhead to project timelines and budgets.

The AI Risk Management Framework (AI RMF), proposed by NIST [13], offers a structured and voluntary methodology designed to help organisations identify, evaluate, and mitigate risks related to AI systems. The AI RMF is acknowledged as a fundamental standard that advocates for the trustworthy and responsible deployment of AI technologies.

Cloud FinOps principles offer a compass for infrastructure cost optimisation including a deeper understanding of utilisation efficiency metrics and a classification of infrastructure based on model archetypes demand. Dimensions such as rightsizing, spot vs. reserved instances, and idle resource trimming are key practices in cloud cost optimisation. FinOps practices and scenario analysis should be fully integrated into AI governance.

3.2.8 Costs of Agentic AI

A number of metrics are emerging to calculate the costs of Agentic AI systems, particularly within the context of large language models (LLMs) and multi-agent systems:

The Agent Cost per Completed Task (ACCT) measures the resources consumed to successfully finish a defined piece of work (task) by an AI agent. It divides the total cost (tokens, compute, API costs) in a given time period by number of tasks completed in the same period.

The Context Memory Optimisation Score (CMOS) is an efficiency metric for how effectively an AI agent manages its context and memory, which are critical and costly resources in LLM-based agents. The goal is to measure the agent's ability to maintain high performance and reliability while minimizing the resource-intensive context window size (i.e., tokens).

The Effective Context Utilisation (ECU) is a metric focused on the quality and relevance of the information an agent uses from its context window to make decisions or take action. The goal is to quantify how well the agent is focusing on the relevant information in its prompt (the "signal") versus getting distracted by irrelevant information (the "noise"). It measures the total tokens influencing a correct action divided by the total tokens in a context window.

In the implementation of Agentic AI Data management, data governance, pricing models and business change management are likely to be the largest cost drivers.

3.2.9 Survey Insights: TCO by AI Lifecycle and Cost types

Based on structured survey responses, two AI use cases at different stages in the lifecycle were provided, *Joint AI Solution 1* (scaled) and *Joint AI Solution 2* (pre-scale). Cost transparency is powerful for predictions, cost controls and for future optimisation of AI use case development process.

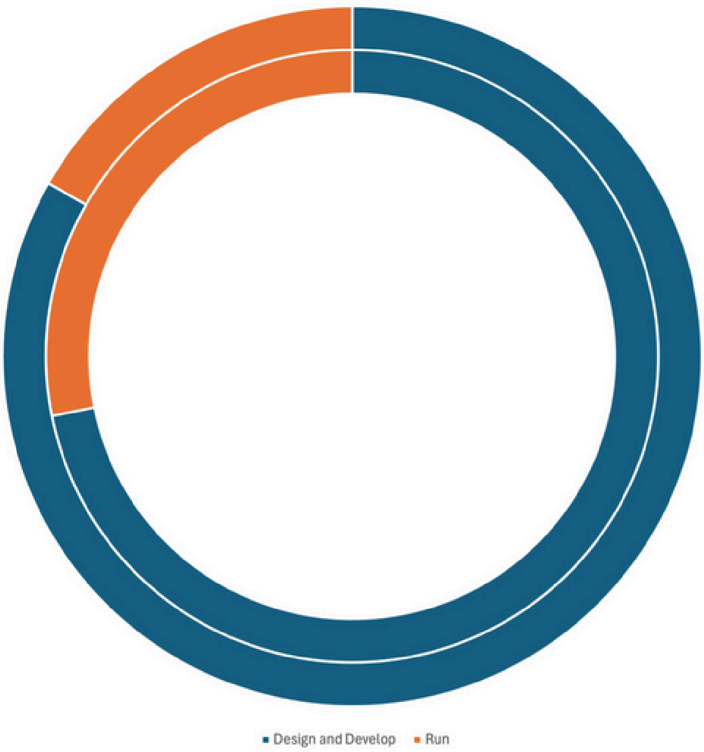


Figure 2: Design and Develop costs vs. Run

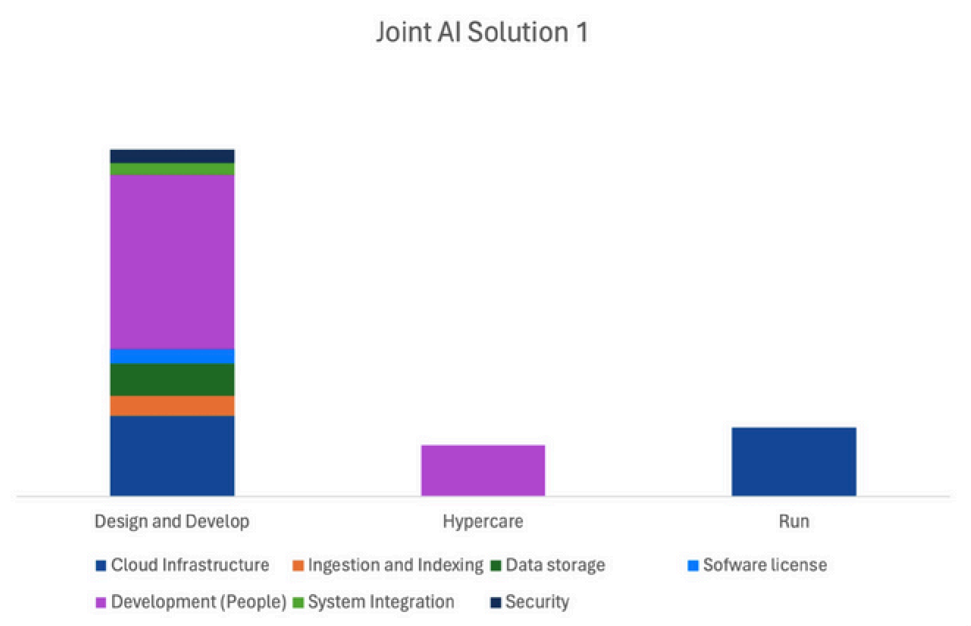


Figure 3: Joint AI Solution 1: TCO by lifecycle (structured survey data; excludes hidden costs such as coordination and stakeholder management).

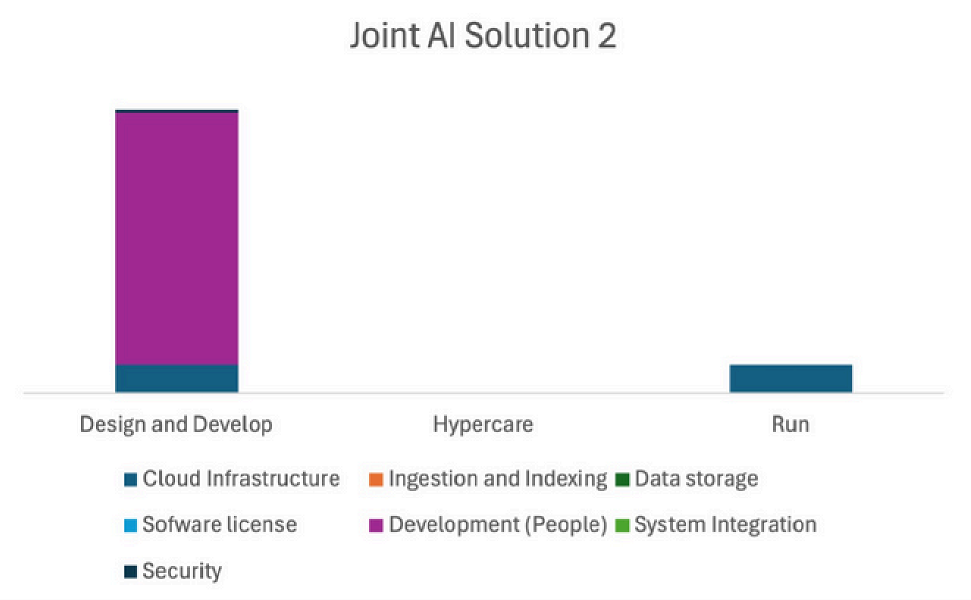


Figure 4: Joint AI Solution 2: TCO by lifecycle

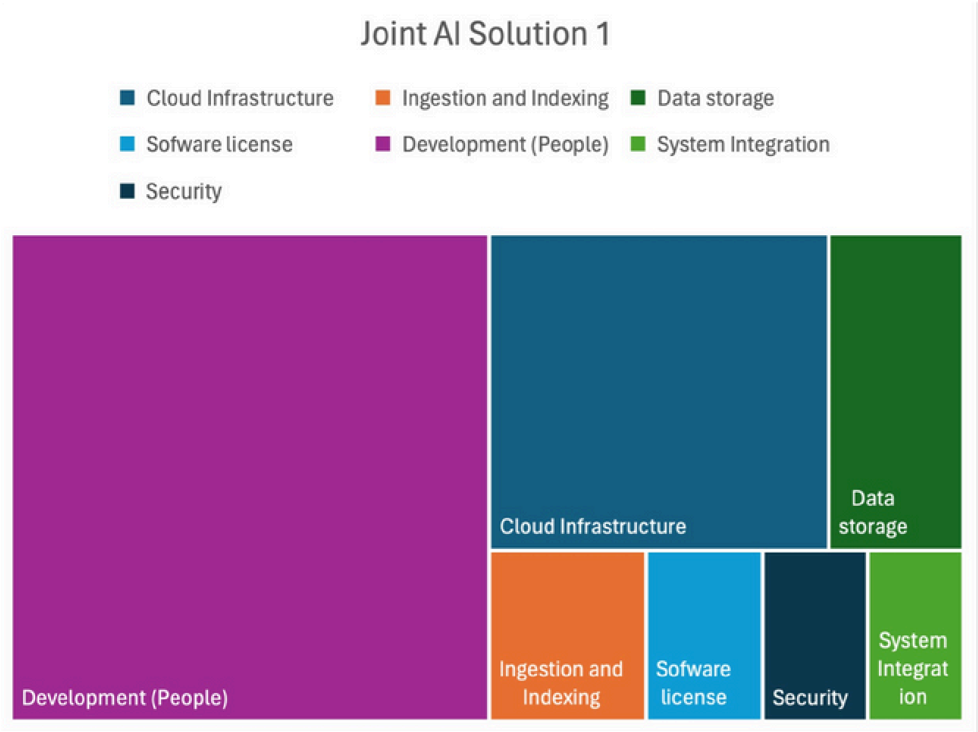


Figure 5: Joint AI Solution 1: TCO by cost type

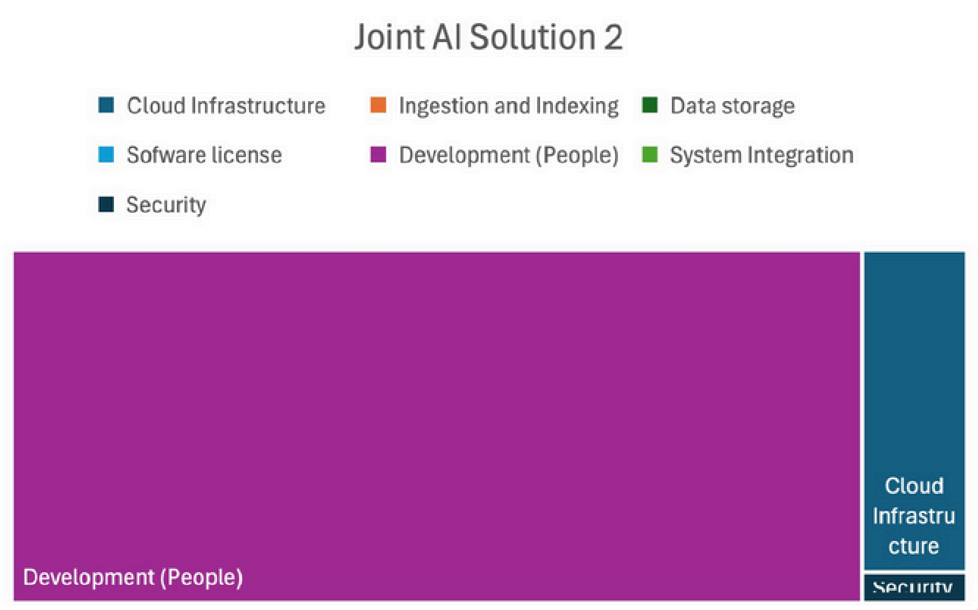


Figure 6: Joint AI Solution 2: TCO by cost type

Across the seven cost pillars, survey evidence highlighted that personnel, architecture, and coordination are dominant cost drivers in the early phases, while integration complexity and governance add substantial costs as projects scale. These findings validate the TCO framework by grounding it in the lived experience of UNICC project teams, showing that the categories are not only theoretically robust but also practically observable in real-world implementations.

In addition to the structured cost categories shown, the survey responses also highlighted a range of hidden or less-visible costs. These included coordination and documentation, stakeholder management, and early-stage overheads associated with the maturity of the AI Hub. Such factors were noted as significantly influencing the Design and Develop phases, and demonstrate why both quantitative and qualitative perspectives are required to fully capture the Total Cost of Ownership across the AI lifecycle.

To synthesise the expanded scope of resources analysed across the TCO, ROI, and SDG layers, Figure 7 illustrates the extended AI Value Chain Matrix. It explicitly adds People and Energy as distinct categories, connecting each to cost, value, and impact metrics across the Frugal AI framework.

AI Value Chain	**TCO — Cost**	**ROI — Value**	**SDG / Ethics — Impact**
People	Staff time, training, coordination, change management	Productivity, service quality, knowledge retention	Inclusion, accessibility, SDG 8
Energy	Power, cooling, energy mix, carbon footprint	Efficiency gains, utilisation optimisation	SDG 7 (Clean Energy), SDG 13 (Climate Action)
Compute	Cloud/on-prem, chipsets, networking	Throughput, latency, resource utilisation	Resource efficiency, SDG 12
Data	Collection, storage, governance, deletion	Insight quality, bias mitigation, reuse	SDG 9 (Innovation), SDG 16 (Institutions)
Governance	Compliance, security, audits, risk management, ethics	Trust, risk reduction, time-to-approval	Responsible AI, SDG 16

Figure 7: AI Value Chain Matrix linking resource categories to cost (TCO), value (ROI), and impact (SDG / Ethics) dimensions within the Frugal AI framework.

3.3 Level 1(C): Fair and Transparent Cost Chargeback Models

As shown in Figure 7, the inclusion of *People* and *Energy* as distinct resources enables a more holistic accounting of AI's cost and value drivers.

A credible cost chargeback model is transparent, equitable, and easily understandable by key stakeholders and partner agencies. The UNICC already tracks key usage metrics and aims to charge back applicable lifecycle costs to such partner agencies. The following model formalises and builds on the current practices of the UNICC. The core principle is to differentiate the costs of developing a new model (amortised over 3 years) against the applicable costs of serving the AI model.

Another source of inspiration is an article by IBM [10] that highlights the essential function accurate allocation of overhead costs plays in understanding and managing TCO within organisations. It delineates the principles of source-drive-target-offset, crucial elements in the distribution of expenses across various departments, products, or projects, thereby enhancing the precision of financial reporting and elucidating actual profitability. The viewpoint enriches the TCO methodology by providing a framework for attribution of indirect costs, which are often neglected by traditional models. By allowing organisations to allocate overhead in a strategic manner, IBM's approach fortifies decision-making about investments, utilisation of resources, and evaluations of products or projects. It emphasises that grasping TCO involves more than aggregating costs; it requires distributing them in a way that most accurately represents value and consumption.

The authors have analysed the responses to a detailed questionnaire compiled by UNICC that included a high level of detail for two use cases Joint AI Solution 1 and 2. In particular, the authors have analysed the detailed costs, aggregated the primary cost drivers, by lifecycle, to propose a transparent cost-chargeback model as follows:

1. *Joint AI Solution 1* shows that, over the first 3 years, 85% of the cost is associated with the design and development of the model (mostly attributed to personnel costs) and the operating costs are 15%. It is also determined that the design and development costs can be amortised over 3 years, whilst the assumption is that the operating cost should be charged back annually.
2. *Joint AI Solution 2* shows a similar pattern, however it is not yet scaled.

In order to establish a logical cost chargeback model, this paper differentiates the **main beneficiary (i.e., the originating function of the AI model) and the users (i.e., the partner agencies).** The authors propose a tiered model for the cost chargeback model i.e. adopting a small/medium/large weight-based system to ensure the use of AI is reasonable to the partner agency.

This paper recommends that the design and development cost be charged back over 3 years to the Main Beneficiary and Users. The paper also recommends that the operational running cost is charged as per Table 1 to the Users as per utilisation or the tiered model.

Phase	Main Beneficiary	Users
Design and Develop	20% to 100%	Tiered Model
Run and Improve	0%	Consumption based + tiered

Table 1: Summary Cost Chargeback Model

3.3.1 Design and Development Costs

The Total Design and Development Cost (TDDC) corresponds to the initial cost component I in Equation 1.2. Unlike recurring operational and maintenance costs, I is a one-time expenditure that must be allocated across users. We compute TDDC as follows:

$$TDDC = \gamma \times \text{Main Beneficiary}_{User} + n \times \text{Allocation Key}_{User}$$

where γ is the coefficient and n is the number of total users. This formulation reflects that design and development costs are borne upfront but distributed over users according to agreed allocation mechanisms.

Decoding the equation: The coefficient, γ is the proportion of the total cost covered by the main originating business function (*e.g., Joint AI Solution 1*). For example, a coefficient of 20% assigns more weight to the primary beneficiary. The *Allocation Key* then distributes the remaining costs among other users. These costs may be charged back in tiers (see Section 1.4.2) or amortised across the expected lifecycle of the AI solution. This ensures consistency with the framework in Equation 3.2, where I is treated as a one-time initial cost distinct from recurring O_t and M_t.

The following flexible approach is suggested:

- **Primary (Recommended) approach:** Tiered Allocation, wherein users should be grouped into tiers according to existing UNICC segmentation based on their size or anticipated benefit. A tiered-pricing model should be established for each tier at the onset. This model offers predictability in budgeting for Users.
- **Alternate approach:** Equal Share Allocation wherein for foundational capabilities that provide equal value to the Users, shared costs are divided equally.

Amortising Initial Investments: For new multi-user platforms or applications, the significant one-time costs of the Design and Pilot phases should be treated as a shared Capital Expenditure. These costs should be amortised over the expected lifecycle of the applications (e.g., a 3-year period); and included in the shared costs that will be charged back quarterly or annually.

It is highly recommended to create a central CapEx "fund" to support participating Users for AI deployment regardless of the ability to fund such AI developments. In such scenarios, the Main Beneficiary could potentially undertake 100% costs to drive faster adoption.

3.3.2 Operating Costs

Operating or Running costs of the model should be allocated by usage (Pay as you go), and by tiered model allocation for fixed costs non-consumption based (for example, for further AI model developments).

Example: *UNICC's* *Joint AI Solution 1*

Phase Number	Scope	Budget (in USD)
Phase 1	Planning/Workshops	14.9
Phase 2	Development	59.2
Phase 3	Hypercare	10.9
Run Costs	Infrastructure: GPT, RAG, index, DB	15

Table 2: Summary Cost: Anonymised by lifecycle stage, and anonymised to USD 100

Note. *From the table, we observe that for phases 1,2 and 3, the total cost = USD 85 (to be amortised in 3 years). Total Run Cost = USD 15 (annual) and Number of Users (agencies) = n.*

Example	Main Beneficiary	Users
Design and Develop	20% over 3 years	Equal allocation/3 years/n
Run and Improve	0%	Equal allocation/n

Table 3: Summary Cost Chargeback Model with cost distribution

Bringing to life the charge-back model with an example, the main beneficiary would pay as charge back USD 5.67 for 3 years and with a simplified equal allocation model users would pay USD 2.89 (USD 1.74 + USD 1.15) for 3 years and USD 1.15 thereafter, assuming Run and improve costs remain constant. Alternatively, Run and Improve costs could be assigned per consumption + a portion of non-consumption-based costs allocated either on a tiered or equal allocation model. A minimum number of agreed user/base consumption could be agreed a basis.

Years 1 to 3	Main Beneficiary (USD)	Users (USD)
Design and Develop	$85 \times \frac{0.20}{3} = 5.67$	$\frac{1}{3} \times \left[\frac{85 \times 0.8}{13}\right] = 1.74$
Run and Improve	–	$\frac{15}{13} = 1.15$
Year 4 onwards	**Main Beneficiary (USD)**	**Users (USD)**
Design and Develop	–	–
Run and Improve	–	$\frac{15}{13} = 1.15$

Table 4: Summary Cost Chargeback Model with cost distribution

3.4 Level 1(D): Measuring Frugal AI at Level 1

This paper proposes to focus on Frugal AI by ensuring that the AI models deployed within UNICC are designed, selected, and maintained for maximum efficiency for both cost and environmental impact. Efficiency in this context refers to computational performance as well as to how resource usage translates into measurable value and alignment with broader UN sustainability goals.

Machine learning models perform a variety of tasks, including language modelling, classification, regression, and multimodal inference. Given the ubiquity of large language models (LLMs) within the UN ecosystem, the baseline metrics use tokens [19] as the primary unit; however, these can be generalised to samples or other relevant units depending on model archetype (e.g., seat-based licencing costs for ML Ops, images for vision models, audio segments for speech models).

3.4.1 Core Efficiency Metrics

The efficiency of these models can be defined through metrics that link computational demand, financial cost, and environmental impact:

- USD/token – cost measurement per token (for models hosted in the cloud). This is often use as metric are more readily available. It is worth noting that this does not measure the cost efficiency of a model per se, as it compounds other factors like energy price, required inference speed or throughput, commercial markup, etc. At model level, the following metrics may serve as a more precise measure of efficiency.

- Tokens/joule – number of processed tokens per unit of energy consumed.

- Tokens/kgCO_2eq – tokens processed per kilogram of equivalent emissions of CO_2.

- Latency and throughput – time to first token, tokens/second; these can be measured at the user's device or system edge for hosted models.

Independent comparative results by Tomlinson et al. (2024) [18] suggest these energy and emissions metrics are material, with AI tasks in writing/illustration emitting orders-of-magnitude less CO_2eq than human equivalents. For UNICC-hosted solutions in particular, energy metrics can be measured using wall-mounted power meters or software-based tools [4]. For hosted services, only financial cost per token may be directly available, requiring estimation models to infer energy usage.

It is important to note that model size (number of parameters) and FLOPs are unreliable proxies of energy cost and should not be used as sole indicators of efficiency. For example, a mixture of experts (MoE) language model might require as much memory as a non-MoE model, with the same number of parameters but far less energy per token [1]. Instead, empirical measurements, such as those listed above, should be prioritised.

3.4.2 Operational Guidance for Measuring Frugal AI

To ensure that efficiency measurement is actionable and comparable across the AI portfolio, the following operational standards are recommended:

1. **Frequency of measurement** – Collect metrics continuously when possible, with monthly aggregation for reporting and benchmarking.
2. **Ownership** – Assign measurement responsibility to the AI governance function within UNICC, supported by technical leads in each agency.
3. **Benchmarks and thresholds** – Establish baseline performance per model type and set efficiency improvement targets (e.g., a 10% reduction in tokens/joule over 12 months).
4. **Normalisation across model types** - Introduce a normalized efficiency score (NES) that converts tokens, samples, or other task-specific units into a standardised efficiency index, enabling a fair comparison between NLP, vision, and multimodal systems.
5. **Lifecycle context** – Measure efficiency separately for the design, development, and production phases, recognising that the efficiency profiles often change significantly between these stages.

3.4.3 Accounting for Hidden Efficiency Costs

Beyond direct computational usage, indirect factors can impact efficiency and therefore are recommended to be considered. Factors include:

- **Data pipeline inefficiencies,** such as repeated ETL processing for the same datasets.
- **Redundant computation cycles** during inference due to suboptimal batching or caching strategies.
- **Retrieval and indexing overhead** in retrieval-augmented generation (RAG) systems that may offset token-level efficiency gains.

By including the aforementioned hidden costs in efficiency reporting, UNICC can make more informed decisions about model optimisation and retirement.

3.4.4 From Measurement to Decision-Making

Efficiency metrics must directly inform AI portfolio decisions — for example, identifying highcost, low-efficiency models for optimisation, scaling energy-efficient models, or decommissioning resource-heavy legacy systems. Where possible, efficiency improvements must be explicitly linked to Level 3 SDG metrics, such as carbon reduction (**SDG 13**), energy efficiency (**SDG 7**), and resource-use efficiency (**SDG 8**).

By integrating financial, technical, and environmental efficiency metrics into a unified framework, monitored consistently and benchmarked across all AI workloads, UNICC can ensure that Level 1 Frugal AI principles are operationalised, measured, and tied directly to strategic goals. This will not only reduce costs but also reinforce UNICC's commitment to sustainable and also responsible AI deployment.

4. Measuring value

For the purposes of this paper, the authors focus on the financial component of ROI, with social impact and other non-financial benefits addressed separately under Level 3 metrics. This approach ensures a precise, quantifiable foundation for decision making while recognising that the broader value of AI deployments will be captured elsewhere. The authors took inspiration from a recent model focusing on the creation of a Value Score as Nominator and a Cost score as Denominator [12] to implement this approach.

A recent report by Forrester Research [7] offers a strategic framework for managing and enhancing expenditures related to AI. It differentiates between two types of cost levers; direct levers, which refer to the expenses associated with models, data, and infrastructure, and operational levers, which involve governance practices, transformation of business processes, and training of the workforce. The report emphasises that data quality, scope, and management are the most critical elements that affect the cost efficiency of AI models. The framework stresses on the financial discipline necessary for effectively scaling AI initiatives, by redirecting focus from performance outcomes to cost structures and operational sustainability. The AI lifecycle management asserts that cost optimisation transcends mere technical issues and constitutes an organisational capability that requires strategic governance and ongoing process enhancement. This framework aims to reconcile AI innovation with financial responsibility, addressing the significant gap between theoretical cost models and actionable business strategies.

4.1 ROI approach

To calculate the benefits for ROI calculation, we propose a **Value Score**. The Value Score reflects the **potential future benefit** (tangible, such as headcount reduction, or intangible, such as productivity gain) enabled by the AI use case. Productivity gains are often the main focus area of AI use cases, however it needs to be noted that often they are not fully harvested as costs savings or revenue growth.

The **Value Score** can be expressed through multiple drivers:

- **Time savings** - for example, reduction in hours required for a task.
- **Cost avoidance** – preventing future expenses, such as avoiding renewals of licences for replaced systems.
- **Error reduction savings** – financial benefits from fewer operational rework, or service incidents.
- **Potential additional User cost enablement** – the ability of UNICC to serve more users or deliver additional outputs without proportional increases in cost.

Using the use-case example *Joint AI Solution 1* (scaled), the authors propose that:

$$\text{Value Score} = \text{Time Saved} \tag{4.1}$$

where Time Saved could be calculated as: the hours previously spent searching 1.3 million documents for functional queries, minus the time now spent prompting the model and validating results, multiplied by the (average) hourly salary of functional staff, providing a direct financial benefit estimate.

This paper defines ROI as:

$$ROI = \frac{\text{Value Score}}{\text{TCO}} \tag{4.2}$$

where the **Value Score** aggregates all applicable benefit categories above.

This framework enables UNICC to prioritise projects, emphasising strategic value per unit cost rather than focussing only on minimising costs. Refer to Section 1 for the Cost dimensions TCO along the different lifecycle states (design and develop, Hypercare, and BAU).

Improvements such as hallucination detection (reducing the need for validation) or prompt templates (shortening prompt design time) would further increase ROI. To ensure comparability and rigour, these benefits are evaluated against the TCO across lifecycle stages: Design & Develop, Hypercare, and BAU.

Linking the AI use case with financial processes and having a thorough baseline measurement before the rollout of the AI use case and after, allows to validate over time [1] if the initial estimates of the **Value Score** have translated into realized cost savings.

4.1.1 Operational Guidance for Measuring Frugal AI

The **Agent Value Multiple (AVM)** is a **Return on Investment (ROI)** metric for an AI agent or multi-agent system. It is a ratio that quantifies the value generated by the agent relative to its cost. The goal is to move beyond simple cost savings (like replacing a human) and measure the **transformative business outcome** driven by the agent.

The AVM measures the Total value generated (including new revenue, efficiency gains, risk reduction) divided by the Total Cost of Agent deployment (which includes Compute, infrastructure, human oversight).

4.1.2 Operationalising ROI Measurement

To make ROI measurement consistent and actionable, the authors recommend:

1. **Frequency** – ROI to be calculated at key milestones (e.g., post-pilot, 6 months after go-live, annually).
2. **Ownership** – The AI governance function at UNICC to lead the ROI tracking based on pre-identified KPIs, in collaboration with User-specific stakeholders or Champions.
3. **Scenario and Sensitivity Analysis** – ROI should be modelled under low, medium, and high adoption scenarios, with sensitivity testing for key assumptions (e.g., wage rates, adoption rates, model accuracy).
4. **ROI Tracking** – Capture Actual vs. Target ROI over the lifecycle of Use cases and at portfolio level
5. **Portfolio-Level ROI** – Aggregate individual project ROI to evaluate the total value of the portfolio and identify the interdependencies where an AI solution enables another.

4.1.3 Empirical Modelling

The authors validated this approach with a simulated [2] multiple linear regression model, quantifying the relationship between ROI and its financial drivers (development cost, operational cost, maintenance cost, and generated value). In this model, the data point for TCO for each row is calculated on the basis of the following formula, which is a modification of Equation 3.1:

$$\text{TCO} = \text{Development Cost} + \text{Operational Cost} + \text{Maintenance Cost} \tag{4.3}$$

This simulated dataset uses uniform distributions [3] (datapoints are uniformly distributed within a specific range of values) for development, operational, and maintenance costs. Each cost value is equally likely and independent from the others. Altogether, the lack of programmed correlation means component costs and revenues are mutually independent.

[1] It's important to note that there is a time differential: while costs are incurred as of Year 1, benefits included in the Value Score may start to be realised as cost savings after Year 3.

[2] The data in this analysis is synthetically generated using random number functions to simulate a range of development, operation, and maintenance costs as well as revenue gains for 100 hypothetical projects. The rationale behind this approach is further explained in the following subsections.

[3] The uniform distribution is preferred since it assumes that all values within a specified range have an equal probability of occurring, implying no preference for any value over another within that interval.

The model predicting ROI (same formula as Equation 4.2) uses statistically significant parameters including development cost, operation cost, maintenance cost, value (hereby defined as revenue gains plus normally distributed noise to represent natural variability), and certain interaction and higher order terms. shows over 95% explanatory power, with all predictors statistically significant at a high confidence level. Higher-order terms further reveal the importance of accounting for **non-linear cost-value interactions**, such as diminishing returns from incremental investment beyond an optimal point.

Note. *The regression results below and throughout this paper are presented in formatted tables, where each row in the subtable titled 'Coefficients' corresponds to a predictor variable and each column therein shows the estimated coefficients, standard errors, test statistics, and associated pvalues. Coefficients with smaller standard errors relative to their magnitude and significance stars (e.g., ***) indicate more robust and statistically meaningful relationships between the predictor and the response variable, holding other variables constant.*

As shown in the table below, ROI is significantly influenced by Value ($p < 0.001$), Development Cost (both linear and quadratic terms), their interaction i.e. Value × Development Cost, as well as Operational and Maintenance Costs. This highlights the multi-dimensional nature of ROI drivers rather than dependence on a single cost factor.

```
Call:
lm(formula = ROI ~ DevCost + I(DevCost^2) + OpCost + MaintCost +
Value + DevCost:Value, data = data)

Residuals:
Min         1Q       Median     3Q         Max
-0.129804 -0.032419 -0.003469  0.023623  0.242046

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)      7.255e-01  1.131e-01   6.416  5.8e-09 ***
DevCost         -4.802e-03  1.753e-03  -2.740  0.007366 **
I(DevCost^2)     2.811e-05  7.413e-06   3.792  0.000265 ***
OpCost          -7.222e-03  3.482e-04 -20.737  < 2e-16 ***
MaintCost       -8.013e-03  6.943e-04 -11.542  < 2e-16 ***
Value            9.925e-03  3.734e-04  26.578  < 2e-16 ***
DevCost:Value   -3.913e-05  3.534e-06 -11.072  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.05829 on 93 degrees of freedom
Multiple R-squared:  0.9833,    Adjusted R-squared:  0.9822
F-statistic: 912.3 on 6 and 93 DF,  p-value: < 2.2e-16
```

Figure 8: Regression model output showing ROI's relationship with development cost, operational cost, maintenance cost, and value, including quadratic and interaction terms.

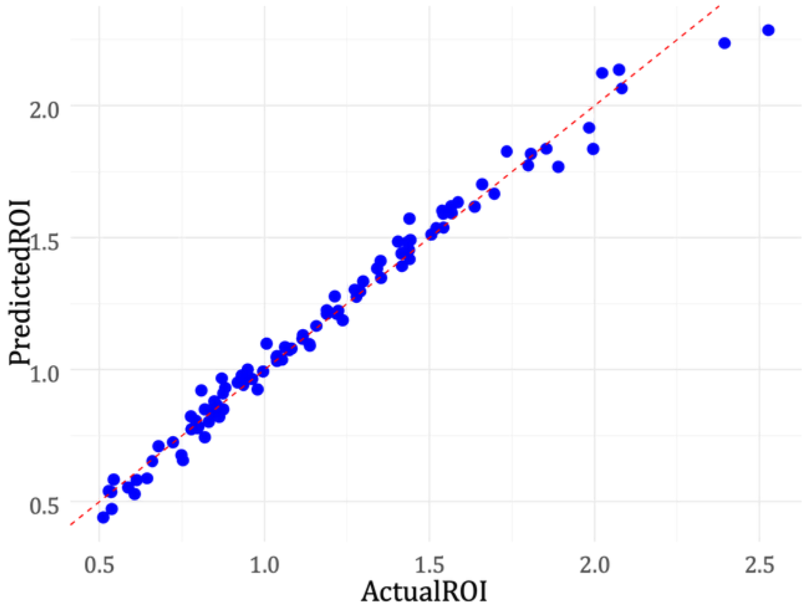


Figure 9: A regression plot showing actual versus predicted ROI for a set of AI projects. The tight alignment around the red dashed diagonal line indicates a strong model fit.

4.1.4 Benchmarking and Continuous Improvement

To maintain credibility, ROI results should be benchmarked against:

- **Industry metrics** for comparable AI use cases.
- **Historical UNICC project data** to track efficiency gains and value delivery improvements.

By embedding ROI measurement into UNICC's AI governance process, with clear benefit categories, lifecycle awareness, sensitivity analysis, and benchmarking, leadership can make confident, data-driven decisions that optimise both financial returns and strategic impact across the multiagency AI portfolio.

To sustain accuracy and organisational learning, UNICC should institutionalise six-monthly reviews of these metrics to refine benchmarks, incorporate new use cases, and share learnings across agencies.

4.2 Modelling AI Project Costs

In this section, the authors briefly explore a multiple linear regression approach. The aim is to examine the relationship between AI *projected* costs by computing a "**Cost Score**" [4] (including two examples: *Joint AI Solution 1 and Joint AI Solution 2*) and Total Cost of Ownership. The objective is to test, with statistical significance, which factors explain "AI project costs", and provide a data-backed evidence base for decision-making when considering frugal AI methods. We gain inspiration from a recent model for Cost-Benefit-ROI that is AI-assisted and specifically designed for the assessment of AI applications [12].

4.2.1 An explanation for the simulated dataset

Recognising a lack of data due to organisations not yet collecting it, time constraints, and the ever-changing market scene, meaning values change rapidly, the authors use a simulated dataset comprising 17 projects, one of which includes the *Joint AI Solution 1 & 2* datapoints (which are not simulated, and directly from the survey data received).

In our attempt at simulating a dataset, and to keep it consistent with the data acquired from the UNICC for *Joint AI Solution 1 & 2* as well as to undergo a broad comparison between AI projects, the total project cost was divided into the subcategories of Acquisition Cost, Operational Cost, and Maintenance Cost, allocating percentages that sum to one hundred, following established practices for generating realistic, compositional data scenarios. The validity of using simulated datasets is well-supported in the literature [20], where simulation facilitates model testing prior to application to real-world data.

A bit more detail about simulating compositional data (data wherein components represent parts of a whole, such as proportions or percentages). This is done by generating values off of a Uniform Distribution and then subsequent normalisation. This approach is commonly used in the literature on compositional data analysis (CoDA), where data represent relative parts of a whole and require special treatment in statistical modelling to avoid multicollinearity — a phenomenon where highly correlated predictors are assessed simultaneously in a regression model [20].

Apart from cost-based factors, our modelling approach also considers data collection mechanisms and timelines. For the former, Type A uncertainty applies if estimates are based on actual measured or observed data over time, whereas Type B uncertainty applies if estimates are based on (external) simulations or assumptions. We use Type A uncertainty for the *Joint AI Solution 1 & 2* survey datapoints, but (randomly yet in an equal proportion) assign that for the rest of the datapoints.

Also, for the time duration, *Joint AI Solution 1*'s is set to 12 months, *Joint AI Solution 2*'s is set to 6 months, whereas for the rest of the datapoints it is (randomly) assigned between 6 months and 24 months (the same timelines as mentioned in our Implementation Roadmap in Section 4.3).

Since (financial) value can be multi-dimensional and hard to measure directly, it is simulated using positive weights on things that increase value and negative weights on things that decrease value. The authors consider multiple variables including different aspects of TCO (namely: acquisition cost, operational cost and maintenance cost), time duration, data collection mechanisms and (empirical) risk. The objective is to identify the factors that most significantly influence project costs, thereby informing strategies for frugal AI deployment. Therefore, for costs, the authors suggest a weighted-scoring model. These methods are widely used to quantify and prioritise projects based on multiple criteria, combining both quantitative and qualitative factors with corresponding weights to generate an overall score that reflects project costs [9].

---

[4] This is not to be confused with the term "Value Score" in the above subsection. Herein, "Cost Score" refers to the (financial) value derived from the AI model.

The authors model using the following formula, obtained using 'handwavy' approximations for fitting multiple linear regression:

$$\begin{aligned}\text{Cost Score} = {} & 0.5 \times \text{Acquisition Cost} - 0.3 \times \text{Operational Cost} \\ & + 0.4 \times \text{Maintenance Cost} + \text{Project Type} + 2 \times \text{Timeline} + \varepsilon\end{aligned} \tag{4.4}$$

where

$$\text{Project Type} = \begin{cases} +20, & \text{if Type A} \\ -20, & \text{if Type B} \end{cases}$$

and the ε term refers to normally distributed random noise (a term that incorporates random variability) such that its mean, $\mu = 0$ and its standard deviation, $\sigma = 16$.

Equation 4.4 takes into consideration how empirical studies suggest long-term acquisition-related spending strongly correlates with project costs, how lower operational costs reduce net project costs, and how sustained investments in maintenance can extend asset life. Also, longer projects imply larger scopes and thus warrant higher investments.

4.2.2 The Model and its Interpretation

```
Call:
lm(formula = CostScore ~ OperationalCost + ProjectType +
TimelineMonths, data = data)

Residuals:
 Min       1Q      Median   3Q       Max
-22.856 -15.159  -1.975  18.037   29.252

Coefficients:
                Estimate   Std. Error  t value  Pr(>|t|)
(Intercept)      67.8288   17.5169      3.872  0.00192 **
OperationalCost  -0.9233    0.2924     -3.157  0.00757 **
ProjectTypeB    -43.5048   10.5201     -4.135  0.00117 **
TimelineMonths    2.4654    0.9894      2.492  0.02700 *

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 19.67 on 13 degrees of freedom
Multiple R-squared:  0.7292,    Adjusted R-squared:  0.6666
F-statistic: 11.67 on 3 and 13 DF,  p-value: 0.0005461
```

Figure 10: Regression model parameters as provided by R, estimating project costs using operational cost, project type, and timeline as predictors.

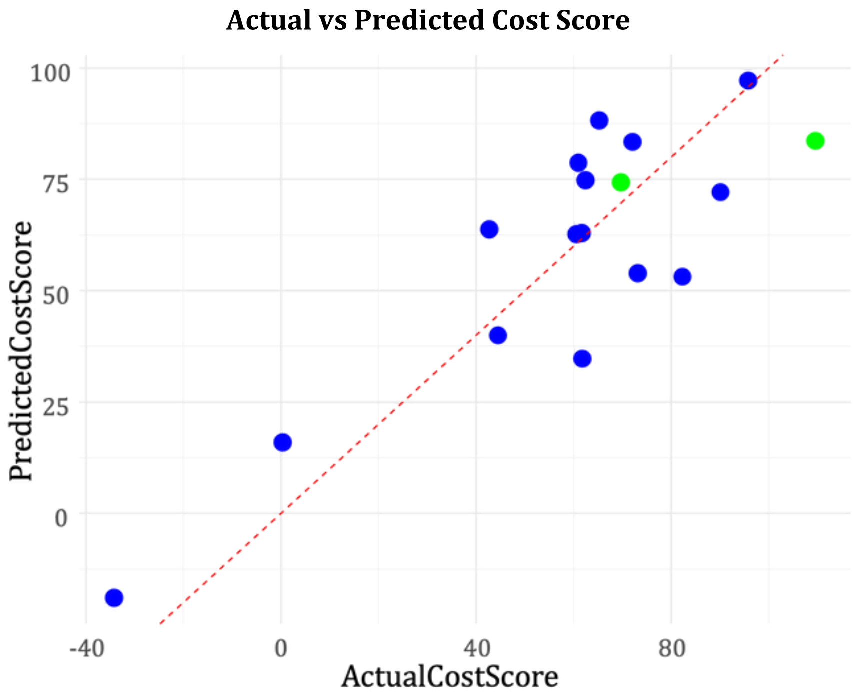


Figure 11: A regression plot of the actual and predicted AI project cost scores, based on the data in Appendix. The strong clustering around the 45-degree line indicates high predictive accuracy.

The multiple linear regression mode [5] exhibits strong explanatory power, with a highly significant overall F-statistic (p = 0.00055). Residual diagnostics indicate moderate prediction error (Adjusted R-squared = 0.6666). These results indicate that the selected independent variables collectively provide a statistically significant prediction of the AI project Cost Score. As shown in the figure above, Operational Cost [6] and project type are highly significant predictors of project costs (p < 0.01), while timeline months is marginally significant (p < 0.05).

A visual comparison of predicted versus observed cost scores demonstrates a strong alignment along the ideal prediction line, reinforcing the model's validity and predeictive accuracy as a tool to support data-driven decision making in the evaluation and prioritisation of AI projects.

A few observations:

- The green points represent Joint AI Solution 1 & 2, which appear in the high-cost end of the spectrum of our graph due to high acquisition / development costs.
- The presence of 2 extreme outliers, one of which has a 'negative' AI project cost, due to high operational cost but significantly lower acquisition cost.

[5] This is a simplified model and that certain non-significant parameters were removed by Backward Elimination technique. In regression analysis, Backward elimination is a stepwise regression technique that starts with all candidate variables (and their interactions) and iteratively removes the least significant predictor based on a chosen criterion, which in our case was the AIC.
[6] This model does not consider all 3 Cost terms to ensure no multicollinearity [20] in the model, since the sum of all costs is considered to be USD 100.

4.2.3 From TCO to ROI to SDG Impact

For non-technical stakeholders, it is important to see three levels of the framework at high level. At its simplest, the framework can be viewed as a flow:

- **Level 1 (TCO):** captures the full cost of design, development, and operation.
- **Level 2 (ROI):** builds on TCO by comparing costs with financial or efficiency returns.
- **Level 3 (SDG Impact):** extends from ROI to include societal and environmental outcomes aligned with the SDGs.

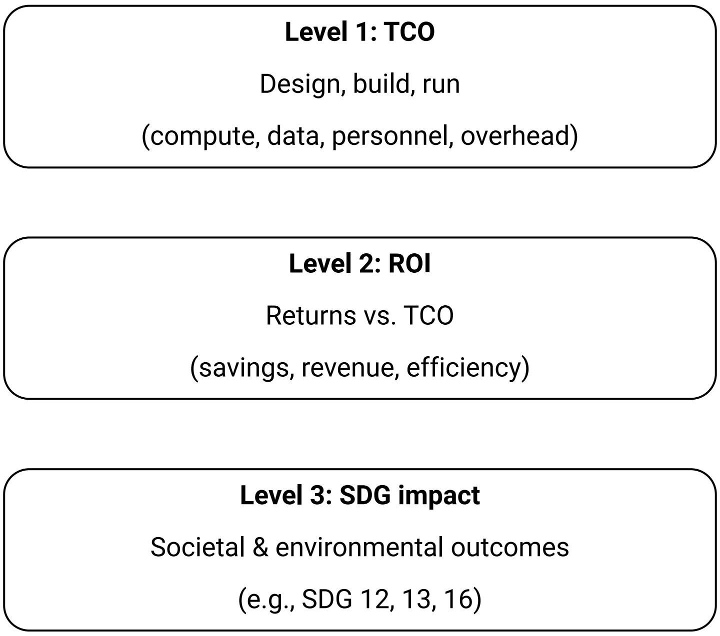


Figure 12: From TCO to ROI to SDG impact: levels 1–3 (vertical layout).

4.2.3.1 Worked Example

Consider a UNICC pilot project using cloud-based AI for document classification (numbers are illustrative):

- **TCO (Level 1):** annualised compute and storage costs are $120k, personnel costs $80k, and data lifecycle costs $50k, giving a total cost of ownership of $250k.
- **ROI (Level 2):** automation saves 6,000 staff hours annually, valued at $300k, generating a net ROI of 20% compared to TCO.
- **SDG Impact (Level 3):** the solution reduces paper use by 1 million pages per year (SDG 12: Responsible Consumption), cuts estimated carbon emissions by 40 tonnes (SDG 13: Climate Action), and improves access to information for staff in multiple regions (SDG 16: Institutions).

This example demonstrates how cost transparency underpins ROI calculations, which in turn create a foundation for measuring broader SDG outcomes. By following this flow, organisations such as UNICC can align AI adoption with both financial and societal objectives.

4.2.4 Possible Extensions

We end this section on modelling costs by understanding that such types of regression models are introductory and may be extended further.

For instance, a recent paper [15] presents a systematic review of AI methodologies utilised in project cost estimation. Specifically, it notes the predictive accuracy of different models. The study reveals that deep learning models attain high accuracy rates (85–90%), surpassing machine learning methods (75–80%) and traditional regression models (70–80%). The research underscores the appropriateness of AI, especially for handling complex, high-dimensional project data. It illustrates how the ability of AI to identify non-linear patterns and interactions enhances forecasting accuracy, leading to improved resource allocation and risk management in projects.

From an advancement in technological standpoint or in terms of what the model is expected to perform, such as modelling project value for machine vision projects, the authors take inspiration from a paper on estimating economic costs of computer vision system using deep learning [17]. This presents a predictive framework that connects the scaling laws of deep learning with the economics of IT infrastructure, aiming to estimate the costs necessary to meet specific accuracy targets in computer vision applications. By merging technical performance metrics with cost modelling, the gap between AI research and its economic viability in practical implementation is addressed. The model offers a quantitative method to predict expenses as the models increase in complexity and data demands. In contrast to general cost frameworks, it specifically models how enhancements in accuracy influence resource utilisation, allowing organisations to make better informed decisions regarding the tradeoffs between performance objectives and financial limitations.

5. United Nations Sustainable Development Goals and Frugal AI

At this level, the framework extends measurement beyond financial return to societal and environmental outcomes, linking the efficiency and value metrics from Level 2 to their broader implications for sustainability, inclusion, and ethical AI use. The following sections focus on the social impact and other non-financial benefits of AI adoption within the Frugal AI framework. We begin by highlighting the impacts of sustainability.

5.1 Studies in the Field of Sustainable AI

This section provides an overview of selected studies and organisational-level research in the emerging field of Sustainable AI, aimed at aligning AI development with environmental and ethical sustainability goals. These examples illustrate how AI-driven indicators can be explicitly mapped to the Sustainable Development Goals (SDGs), offering practical tools for organisations aiming for measurable impact **[19]**.

- The AI for Good initiative has standardised metrics designed specifically to assess AI's contribution to sustainability and social impact [11].

- The OECD Framework on AI and Sustainable Development outlines comprehensive approaches for linking AI adoption clearly to individual SDG targets [14].

- Another study published in Nature Communications in 2020, provides comparative analysis and a detailed evaluation of AI applications across the SDGs [21].

- The Digital Public Goods Alliance (DPGA) offers metrics and case studies focused on the adoption and impact of open, inclusive, and accessible digital technologies, emphasising digital equity and sustainability [5].

- The Swiss Sustainable Investment Market Study, [16], presents a detailed examination of the sustainable investment environment in Switzerland. The report indicates a persistent increase in sustainability-related conclusions of 2024, a 13% rise compared to the previous year, highlighting a robust recovery in spite of global challenges. The report focusses on investor motivations, implying that sustainable investment strategies are increasingly incorporating considerations of long-term cost efficiency and value governance. The advancing maturity of the market suggests that decision makers are taking into account wider, lifecycle-orientated consequences aligning with the fundamental tenets of TCO analysis. Although TCO is not measured directly, the study supported by the SSF offers significant contextual clues that indicate the implicit importance of understanding the sustainability contexts in decision making processes.

- While discussions of AI often emphasise its environmental costs, recent evidence provides an important counterpoint. Tomlinson et al. (2024) [18] show that AI systems can emit between 130–1500× **less CO_2e per page of text** and 310–2900× **less CO_2e per image** than humans performing equivalent tasks. Their analysis explicitly links these results to **SDG 12** (Responsible Consumption and Production) and **SDG 13** (Climate Action), reinforcing why the authors use Level 3 metrics including tokens/joule, tokens/kgCO_2e, and emissions reduction targets as central to evaluating Frugal AI.

5.2 Notes on Specific SDGs

To systematically align Frugal AI metrics with the UN SDGs, the authors have selected specific goals that explicitly link to measurable, actionable outcomes. These are summarised in the table below.

SDG	Relevant Sub-targets	Frugal AI Metrics
SDG 7 Affordable and Clean Energy	**7.2:** Increase substantially the share of renewable energy globally. **7.3:** Double the rate of improvement in energy efficiency globally	• Optimization of renewable energy grid/storage using AI (%) • Energy efficiency (tokens/joule) • Carbon intensity (tokens/kgCO2 eq) • Compute infrastructure efficiency (For example: P.U.E for data centres, Energy per Unit of Work in Cloud)
SDG 8 Decent Work and Economic Growth	**8.4:** Improve global resource efficiency and decouple economic growth from environmental degradation	• % Improvement in resource efficiency due to Frugal AI models • Contribution of AI-enabled processes to green economic activities and sustainable growth
SDG 9 Industry, Innovation and Infrastructure	**9.4:** Upgrade infrastructure to be sustainable with increased resource-use efficiency. **9.5:** Enhance scientific research and technological capacity	• Reduction in infrastructure costs due to Frugal AI (%) (this includes batched compute processes in green hours) • % Reduction in compute/storage required per deployment • Number of Frugal AI solutions deployed in low-resource environments
SDG 10 Reduced Inequalities	**10.2:** Empower inclusion irrespective of age, sex, disability, race, ethnicity, origin, religion, or economic status	• Accessibility improvements (% increase in adoption due to lower barriers such as cost/accessibility) • Effectiveness of cost-tiered chargeback model (adoption rates among small vs large UN agencies) • Multilingual accessibility and inclusion reach (% increase in language coverage and user reach)
SDG 12 Responsible Consumption and Production	**12.2:** Achieve sustainable management and efficient use of natural resources. **12.5:** Substantially reduce waste through prevention, recycling and reuse	• Reduction in redundant computing cycles (%) • Improved lifecycle management of AI hardware and software (reduction in electronic waste %) • Resource efficiency gains from AI deployments (%)
SDG 13 Climate Action	**13.2**: Integrate climate change measures into policies, strategies, and planning	• Reduction in emissions due to Frugal AI (% emissions avoided) • Contribution of AI optimisation to achieving organisational carbon neutrality goals (%)
SDG 16 Peace, Justice and Strong Institutions	**16.6**: Develop effective, accountable and transparent institutions at all levels	• % improvement in decision transparency (e.g., AI-enabled dashboards for reporting and monitoring) • Number of institutions reporting improved accountability and governance through Frugal AI tools
SDG 17 Partnerships for the Goals	**17.16**: Enhance global partnerships for sustainable development	• Number of agencies adopting shared Frugal AI solutions (partnership success) • Efficiency gains through collaborative AI resource pooling (cost savings %)

Table 5: Aligning Frugal AI metrics with the UN SDGs

The authors also reviewed other SDGs such as **SDG 4** (Quality Education) and **SDG 11** (Sustainable Cities and Communities), which were considered of secondary relevance at this stage. This focused SDG framework ensures clear alignment, practical measurability, and actionable guidance consistent with established international benchmarks and frameworks from organisations such as the UN Global Pulse, ITU, OECD, and the Nature Communications study.

To ensure that innovation is both impact-driven and user-centred, this paper recommends that UNICC should incorporate continuous users' feedback mechanisms. The authors also recommend introducing a Net Promoter Score (NPS) framework and usage analytics to monitor agency satisfaction and adoption rates after deployment. These insights should inform decisions about the AI portfolio on whether to scale, enhance, or retire specific AI solutions.

Survey Insight: Early adoption data from UNICC indicates that the tiered chargeback model has supported inclusion by enabling smaller agencies to participate alongside larger ones, and multilingual system requirements have already surfaced as a critical driver of accessibility.

5.3 SDG Score

The authors now proceed to try to quantify the benefits / relationships between the impact driven by the guiding principles of the UNSDGs and the ROI. Calculating a SDG score is complex. Thus, the authors have created a composite metric underpinned by a weighted scoring model that combines reach, number of SDG aspects covered, and energy efficiency.

$$\begin{aligned}\text{Net SDG Score} = {} & w_1 \times (\text{Reach}) + w_2 \times (\text{SDG Aspects}) + w_3 \times (\text{Energy Efficiency}) \\ & + w_4 \times (\text{Equity / Inclusion}) + w_5 \times (\text{Resilience}) - w_6 \times (\text{Harm Index}) +\end{aligned} \tag{5.1}$$

where the **SDG-aligned performance factors** are defined as follows:

- **Reach:** The number of users reached or 100 × (% target coverage).
- **SDG Aspects:** A quantification of the breadth or depth of the solutions' social impact considering the SDG targets addressed.
- **Energy Efficiency:** Metrics such as tokens/joule, compute hours saved, or emissions avoided.
- **Equity/Inclusion:** 100 × (% youth, women, or underserved groups impacted).
- **Resilience:** A factor variable that quantifies the sustainability of the intervention as (pilot = low, systemic change = high).
- **Harm Index:** Penalties for bias, exclusion, misinformation.

Ideally, the weights (w_1, w_2, w_3, w_4, w_5 and w_6) should be determined by stakeholders based on the strategic priorities of the organisation – the only limitation being that the (signed) weights[7] sum up to 1. For instance, an organization prioritizing broad social inclusion might give a higher weight to Reach and SDG Aspects, while one focused on environmental goals might weight Energy Efficiency more heavily. This method is similar to the weighted-scoring model for project costs mentioned in Section 2.2 of the paper, which combines quantitative and qualitative factors with corresponding weights.

[7] The authors assign a negative weight to the Harm Index as a method to penalise.

5.3.1 SDG Score in practice

For an example of the SDG Score in practice, the authors set the weights as follows: w_1 = 0.25, w_2 = 0.20, w_3 = 0.15, w_4 = 0.25, w_5 = 0.10 and w_6 = 0.30. Also, the ε term refers to normally distributed random noise (a term that incorporates random variability) such that its mean, μ = 0 and its standard deviation, σ = 5. Also, the authors cap the **Net SDG Score** between 0 and 100 to ensure the plot and results shows well and for visual cues and understanding.

This simulated dataset which was computed with a similar methodology to that expressed in Section 2.2.1 is in the Appendix. It is visualised as a scatter-plot with the following axis:

- **ROI (X-axis):** For each project, the authors assign a ROI score similar to Equation 4.2.

- **SDG/Frugal AI Score (Y-axis):** For each project, the authors assign an SDG score using the weighted model described in Equation 5.1.

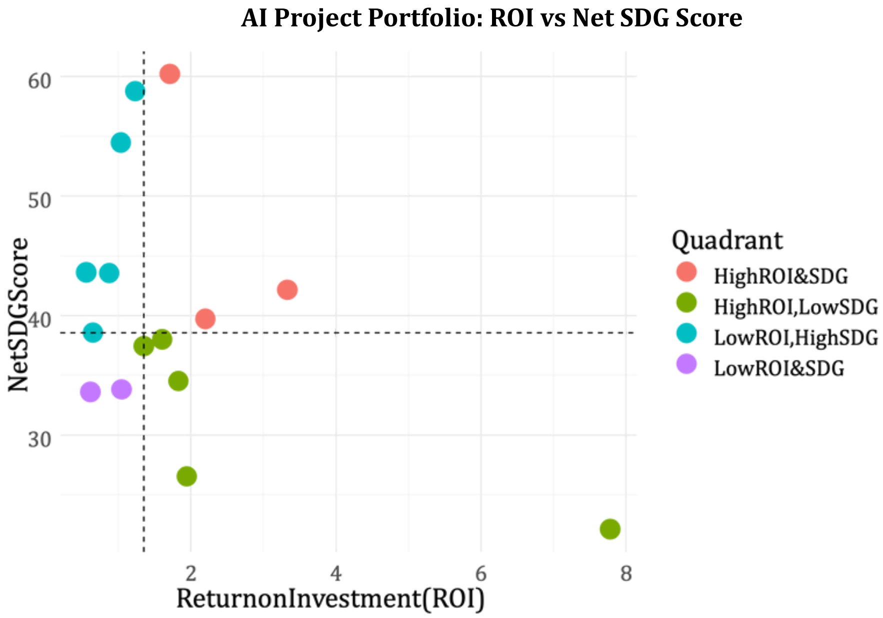


Figure 13: A scatterplot of the simulated dataset on a ROI × SDG Score plane, based on the data in Appendix.

Interpretation: The projects are categorised into quadrants to inform strategic decisions as follows:

- **High ROI, High SDG Score (Top Right):** These are the ideal projects. They deliver both strong financial returns and significant social impact. The organisation should prioritise scaling these solutions.

- **High ROI, Low SDG Score (Bottom Right):** These projects are financially successful but don't align with the UN's broader sustainability goals. They should be considered for optimisation to improve their social impact.

- **Low ROI, High SDG Score (Top Left):** These are socially impactful but financially inefficient projects. The organisation should investigate ways to reduce their TCO or increase their value to make them more sustainable.

- **Low ROI, Low SDG Score (Bottom Left):** These projects should be candidates for a "decommission" or "stop-and-re-evaluate" decision, as they are not meeting either financial or social objectives.

The authors end this section by understanding the basic relationship between ROI and Net SDG Score through the following plot based on a simple linear model between them.

$$\text{Net SDG Score}_i = \beta_0 + \beta_1 \cdot \text{ROI}_i + \varepsilon_i \tag{5.2}$$

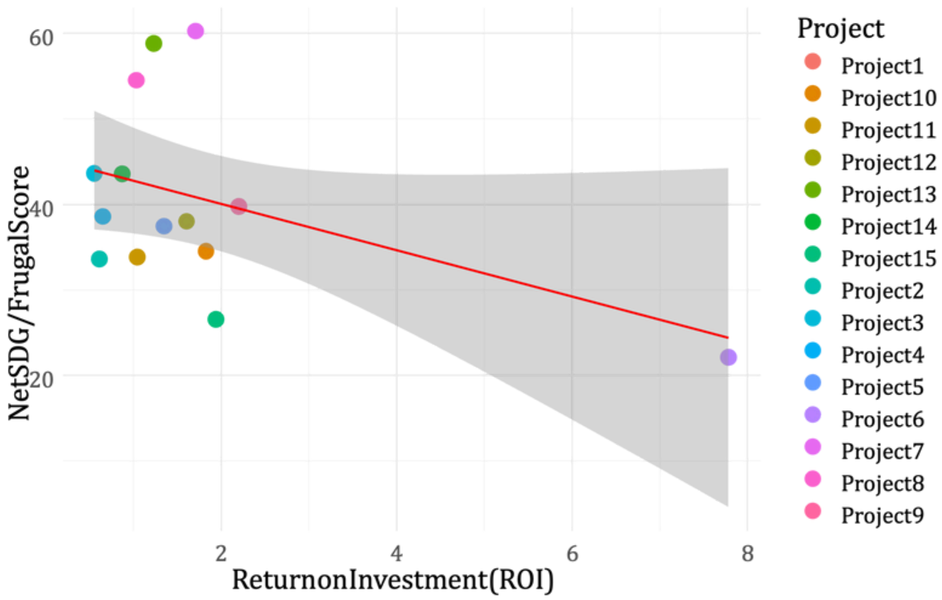


Figure 14: The shaded band around the regression line in the scatter plot represents the 95% confidence interval for the mean Net SDG Score at each level of ROI, based on a simple linear model.

Note. *This interval indicates where the average response is expected to lie 95% of the time. Therefore, it is normal and statistically acceptable for some individual projects to fall outside this shaded area.*

5.3.2 Extensive Statistical Modelling

The authors also proceed with 2 multiple linear models – one for ROI prediction using the Net SDG Score and other components, and another for the Net SDG Score prediction using the ROI and other components. In both models, we begin with the full model with all the components. However, they both give very poor fitting models. Therefore, the models are simplified and certain non-significant parameters are removed by Backward Elimination (as described in Section 2.2.2).

5.3.2.1 The ROI Model

```
Call:
lm(formula = ROI ~ Net_SDG_Score + Energy_Efficiency + Harm_Index,
    data = portfolio)

Residuals:
    Min      1Q  Median      3Q     Max 
-1.4270 -0.6605 -0.4768  0.6375  3.6492 

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)  
(Intercept)        2.07082    1.98670   1.042   0.3196  
Net_SDG_Score     -0.08822    0.03960  -2.228   0.0477 *
Energy_Efficiency  0.03222    0.01852   1.740   0.1097  
Harm_Index         0.05116    0.02643   1.935   0.0790 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.486 on 11 degrees of freedom
Multiple R-squared:  0.4623, Adjusted R-squared:  0.3157 
F-statistic: 3.153 on 3 and 11 DF,  p-value: 0.06852
```

Figure 15: Simplified regression model parameters estimating ROI.

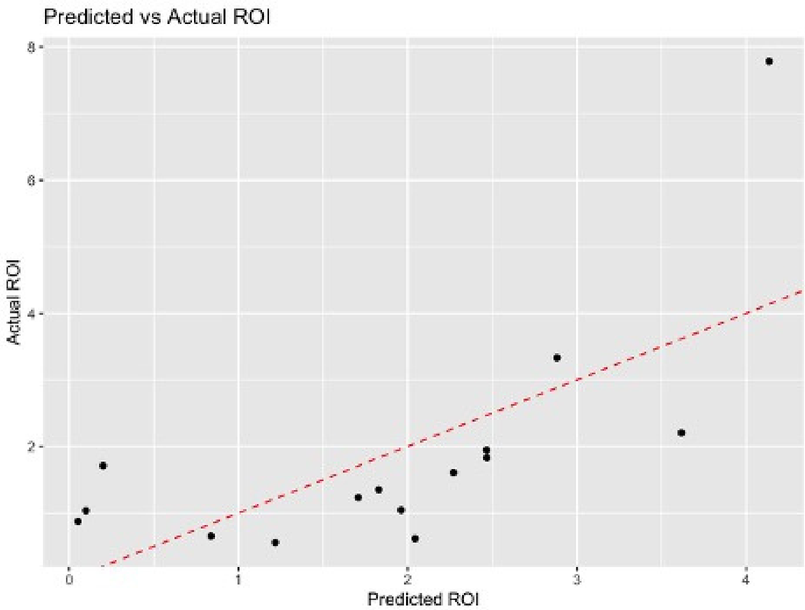


Figure 16: Plot showcasing predicted ROI vs. actual ROI for the above linear model. This shows explicitly that the model does not explain all the variability associated with ROI.

This linear model explains 46% of the variance in ROI, with Net SDG Score emerging as a statistically significant (negative) predictor ($p = 0.048$), and Harm Index shows a marginal association too ($p = 0.079$). Therefore, in order to enhance the variability and add further context to the model, the authorsadd two more ROI explanatory factors – TCO and Value Score to the linear model. This then creates a direct linkage with this anaylsis and the empirical modelling we discuss in Section 2.1.2.

This new model with the 2 added parameters explains 77.5% of the variance in ROI, with TCO and Value Score showing statistically significant effects on ROI ($p < 0.01$), which collaborates with the results from Section 2.1.2. However, herein, in order to add explanatory power, we lose out on the significant statistical dependence of the ROI on the Net SDG Score.

```
Call:
lm(formula = ROI ~ TCO + Value_Score + Net_SDG_Score, data = portfolio)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.08710 -0.61199 -0.03245  0.40929  2.21402 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)   
(Intercept)    3.66082    1.20469   3.039  0.01127 * 
TCO           -0.04643    0.01190  -3.901  0.00247 **
Value_Score    0.03207    0.00852   3.764  0.00314 **
Net_SDG_Score -0.04228    0.02558  -1.653  0.12665   
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9606 on 11 degrees of freedom
Multiple R-squared:  0.7752, Adjusted R-squared:  0.7139 
F-statistic: 12.64 on 3 and 11 DF,  p-value: 0.000691
```

Figure 17: Expanded regression model parameters estimating ROI.

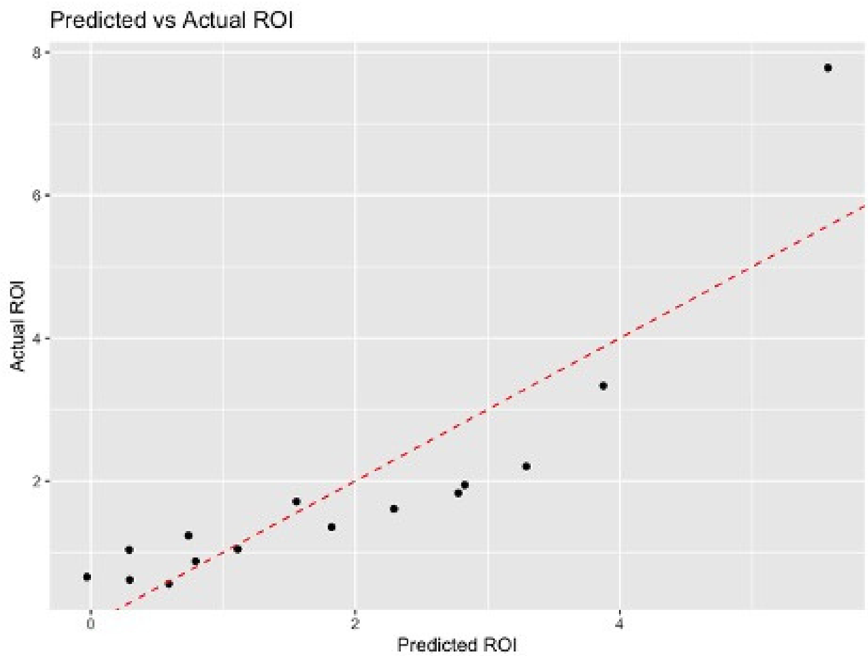


Figure 18: Plot showcasing predicted ROI vs. actual ROI for the above linear model. This shows that the model better explain the variability associated with ROI because of better fit with the 45-degree line. There is quite an evident outlier point, though.

5.3.2.2 The SDG Score Model

```
Call:
lm(formula = Net_SDG_Score ~ Reach + SDG_Aspects + Energy_Efficiency +
    Equity_Inclusion, data = portfolio)

Residuals:
    Min      1Q  Median      3Q     Max
-6.1173 -2.0053 -0.0292  2.5057  3.5466

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)       -13.01086    4.71976  -2.757 0.020245 *
Reach               0.27351    0.03900   7.013 3.66e-05 ***
SDG_Aspects         0.31343    0.03700   8.472 7.11e-06 ***
Energy_Efficiency   0.27250    0.03924   6.945 3.97e-05 ***
Equity_Inclusion    0.15891    0.03231   4.919 0.000606 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.229 on 10 degrees of freedom
Multiple R-squared:  0.9357, Adjusted R-squared:   0.91
F-statistic: 36.38 on 4 and 10 DF,  p-value: 6.245e-06
```

Figure 19: Regression model parameters estimating Net SDG Score.

This linear model explains 93.5% of the variance in the Net SDG Score, with factors about Reach, Energy Efficiency, Equity Inclusion and SDG Aspects showing statistically significant effects on ROI ($p < 0.001$).

This means that the Harm Index may not be statistically significant and therefore we could conduct the analysis without considering it as a part of the SDG score. However, it is essential, from a theoretical standpoint to add a "loss term" in the modelling, alongside the normally distributed noise.

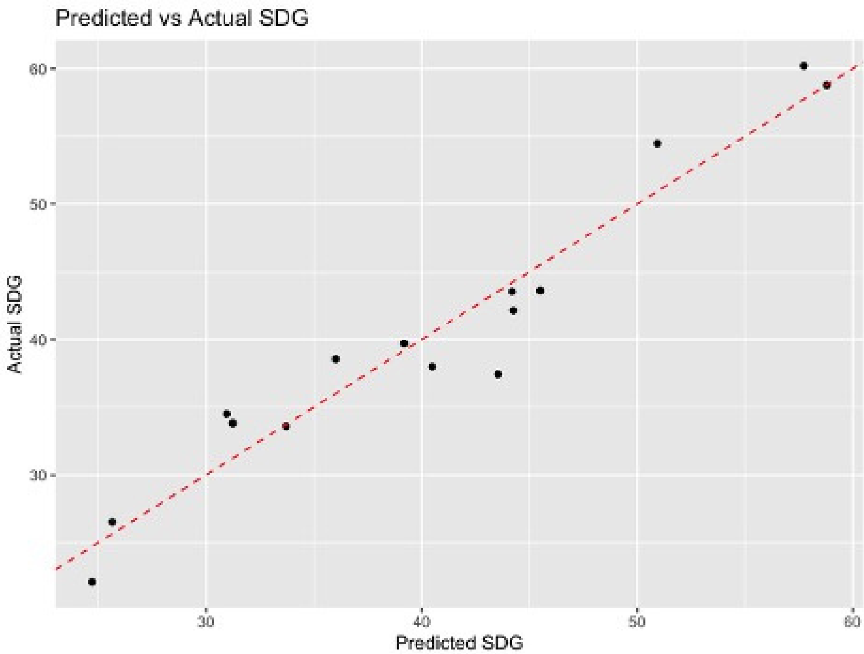


Figure 20: Plot showcasing predicted SDG score vs. actual SDG score from the above linear model. The strong clustering around the 45-degree line indicates high predictive accuracy

5.4 Ethical Costs

In addition to financial, operational, and governance-related costs, organisations should explicitly consider *ethical costs*. These costs are often hidden or under-accounted for, yet they can be decisive for both adoption success and public trust. Ethical costs typically include:

- **Bias and Fairness Audits:** resources devoted to detecting, mitigating, and monitoring bias in data and model outputs.
- **Transparency and Explainability:** costs of implementing interpretability tools, documentation, and communication with stakeholders to ensure responsible use.
- **Safeguards Against Misuse:** investments in processes to reduce risks such as misinformation, exclusion of vulnerable groups, or unsafe deployment contexts.
- **Compliance with Ethical Standards:** alignment with emerging AI ethics frameworks (e.g., UNESCO, OECD, EU AI Act) and institutional review processes.

For example, accessibility adaptations of AI interfaces for visually impaired users represent a quantifiable ethical cost component that may be shared centrally or allocated per project.

Integrating these ethical costs into the Total Cost of Ownership framework ensures a more comprehensive and credible assessment of AI initiatives. By doing so, organisations such as UNICC can anticipate challenges earlier in the lifecycle, strengthen accountability, and align AI adoption with broader societal expectations.

6. Summary Metrics at AI model level

6.1 Metrics suggested for the AI use case portfolio

Level	Metrics
L1: TCO and Frugal AI	1. **TCO (across Design & Develop & Hypercare) and Run elements** • Compute and Infrastructure • Data lifecycle • Model and Software • Personnel • Integration and Orchestration • Governance, Risk, Compliance and Ethics **2. Frugal AI** • USD/token (use with caution, see relevant chapter) • tokens/joule, tokens/kgCO2eq • if latency and throughput are paramount – time to first token and tokens/s.
L2: ROI	$ROI = \frac{\text{Value Score}}{\text{TCO}}$ **ROI factors:** • Frequency • Ownership • Scenario and Sensitivity Analysis • Risk-Adjusted ROI • Lifecycle ROI Tracking • Portfolio-Level ROI **Near Future:** • Agent Value Multiple (AVM)
L3: SDG and Frugal AI	• Energy efficiency (tokens/joule) [SDG 7, SDG 13] • Carbon intensity (tokens/kgCO2eq) [SDG 13] • Compute infrastructure efficiency (Cloud vs. On-premise emissions) [SDG 7, SDG 9, SDG 12] • Reduction in infrastructure costs due to Frugal AI models (%) [SDG 9] • % Reduction in compute/storage resources required per deployment [SDG 9, SDG 12] • Reduction in redundant computing cycles (%) [SDG 12] • Improved lifecycle management of AI hardware and software (reduction in electronic waste %) [SDG 12] • % reduction in emissions due to Frugal AI [SDG 13] • Contribution of AI optimisation to achieving organisational carbon neutrality goals (%) [SDG 13] • Accessibility improvements (% increased adoption due to lower barriers such as cost or device requirements) [SDG 10] • Multilingual accessibility and inclusion reach (% increase in language coverage and diverse user adoption) [SDG 10] • Effectiveness of cost-tiered chargeback model (adoption rates among small vs. large UN agencies) [SDG 10] • % improvement in decision transparency (e.g., AI-enabled dashboards, explainability features adopted) [SDG 16] • Number of institutions reporting enhanced accountability and governance through Frugal AI [SDG 16] • Number of agencies adopting shared Frugal AI models (partnership success indicator) [SDG 17] • Efficiency gains through collaborative AI resource pooling (cost savings %) [SDG 17]
L3: Customer Feedback	• Net Promoter Score (NPS)

Table 6: Level-based Summary

6.2 Operationalising Measurement

The survey responses from UNICC regarding the use cases for J*oint AI Solution 1 & 2* provided a strong foundation for operationalising measurement.

6.2.1 Level 1: TCO and Frugal AI Metrics

1. Cost Tracking & Reporting:

- **Internal timesheets and workshop participation:** Used to track personnel and expertise costs.
- **Support tickets and triage reports:** Used for TCO tracking and reporting.
- **Cloud billing and usage logs:** The primary procedure for tracking TCO.
- **Fixed vs. Variable costs breakdown:** The survey provides specific examples, such as developer time, stakeholder workshops, and queries as variable costs, while Azure licenses and security reviews are fixed.
- **SDA and BCR contracts and approved budget:** The basis for cost structure, maintenance, and support.

2. Operational & Usage Metrics: These are essential for calculating TCO and Frugal AI efficiency metrics like USD/token.

- **Number of queries:** Tracked per agency as part of the cost-sharing model for Joint AI Solution 1 and 2.
- **Number of users:** Tracked per agency for Joint AI Solution 1
- **Number of documents indexed:** A metric for the scale phase of Joint AI Solution 2.
- **Number of workshops and hours logged:** Used to track costs for requirements gathering and solution design.
- **Number of stories completed/defects:** Used for tracking App and UI/UX costs.
- **Emails sent:** A metric for SLA-based support in the scale phase.

6.2.2 Level 2: ROI Metrics

The paper defines the ROI formula as in Equation 4.2 and lists general factors like ”time savings” and ”cost avoidance”. The survey suggests specific, measurable metrics that contribute to the Value Score.
The Value Score based factors include:

- **Productivity KPIs:** Time saved per user, which is a key productivity KPI.
- **Augmented cost efficiency:** A metric that contributes to the overall value of the AI solution.
- **User feedback:** User feedback, specifically ”thumbs up/down,” is mentioned as a proxy for impact.

6.2.3 Level 3: SDG and Customer Feedback Metrics

- **Social Impact and Partnerships**
- **Multilingual inclusion reach:** A metric related to the UN’s mission and its impact on reducing inequalities.
- **User engagement per agency:** A foundational metric for cost-sharing that also indicates adoption and partnership success.

Summary: At **Level 1,** the framework focuses on the **Total Cost of Ownership (TCO)** and **Frugal AI efficiency**. To calculate TCO, organizations must track both fixed and variable costs across the entire AI lifecycle. Operational measurement is facilitated by utilizing existing internal systems, such as cloud billing and usage logs, internal timesheets, workshop participation records, and support tickets. These tools allow for the granular measurement of costs associated with compute, data, and personnel. While USD/token remains a financial metric of inference cost, the Frugal AI metrics complement it with distinct environmental indicators such as energy-mix intensity (gCO_2e/kWh) and compute efficiency (tokens/joule). These indicators allow agencies to separately assess the carbon and resource efficiency of AI workloads, rather than conflating financial and environmental performance.

At **Level 2,** the **Return on Investment (ROI)** metric is defined as the **Value Score** divided by **TCO**. The Value Score quantifies tangible and intangible benefits, such as time savings, cost avoidance, and productivity gains. The paper suggests that these benefits can be measured through specific operational metrics, including time saved per user. To ensure a transparent and credible ROI calculation, the framework recommends that all projects be tracked against pre-identified KPIs at key milestones, with the AI governance function leading this effort.

Level 3 links AI performance to the United Nations **Sustainable Development Goals (SDGs)** to measure social and environmental impact. This involves evaluating the AI portfolio's contribution to goals like energy efficiency (SDG 7) and reduced inequalities (SDG 10). Key metrics for this level may include as an example, multilingual inclusion reach and user engagement per agency, which are used to measure the solution's impact and partnership success. The framework also incorporates a customer-centric dimension by recommending the use of a **Net Promoter Score (NPS)** and other user-feedback mechanisms to assess satisfaction and adoption rates, thereby informing decisions to scale, enhance, or retire AI solutions. It also includes an SDG Score, a quantification of benefits based upon multiple factors encompassing reach, alignment, energy efficiency, inclusion and resilience.

6.3 Implementation Roadmap: A phased approach

Implementation and change management are part of another phase of research in our pipeline. This paper gives a very initial overview which will be further researched and developed in dedicated pilots currently in progress in the Frugal AI Hub.

At high level, a phased approach is likely to ensure a smooth and manageable rollout. For example:

- **Select and agree metrics (Up to 6 Months):**
    - Discuss with relevant stakeholders the proposed approach and select metrics.
    - Define operationalisation routes (how each metric will be measured).
- **Pilot the framework (6 to 18 months):**
    - Apply the refined framework to a smaller AI portfolio of up to 10 initiatives.
    - Collect feedback and refine measurement processes.
- **Scale portfolio-wide (after 18 months):**
    - Automate and implement at portfolio level.
    - Manage the AI portfolio as an organisational asset.
    - Set up AI optimisation processes aligned with budget planning.

6.3.1 Adoption Playbook

To embed the framework into processes, a phased playbook is recommended across procurement, project approvals, and post-implementation reviews.

To start with, scenario planning on Build vs. Buy trade offs at scale, experimenting at portfolio level can bring significant long term benefits.

1. Procurement and Vendor Selection

- *RFP Requirements:* mandate that all proposals include a breakdown of costs aligned with Level 1 TCO pillars (compute, data, personnel) and projections of ROI (Level 2).
- *Weighted Scoring:* apply the framework metrics in vendor evaluation, rewarding energy efficiency (tokens/joule) or lower projected TCO.
- *Supplier Engagement:* require plans for bias, data privacy, and transparency, linked to the Level 3 Harm Index. Assess risks of vendors lock in. Stronger contracts are are key levers for real cost reduction.

2. Project Approval Processes

- *Cost-Benefit Analysis:* new AI projects must submit TCO and ROI analysis, including longterm operational costs.
- *Impact Alignment:* the Level 3 SDG Score serves as a checklist; proposals should show measurable SDG contributions.
- *Prioritisation Matrix:* projects can be plotted on an ROI–SDG Score matrix to visualise priorities.

3. Post-Implementation Portfolio Reviews (PIR)

- *Performance Audits:* conduct regular audits (e.g., quarterly) to measure actual TCO, ROI, and SDG Score.
- *Continuous Optimisation:* use findings to refine deployments (e.g., model compression, efficient data pipelines). Focus on financial efficiency, loss-risk reduction and capital efficiency. Include learnings to improve and accelerate change management, training and integration.
- *Knowledge Sharing:* document successes and failures to inform future projects and refine the framework.

By embedding the framework into procurement, approval, and review cycles, organisations can ensure adoption is systematic rather than ad hoc, and that financial and social impacts are measured consistently.

6.4 Further considerations

The measurement of TCO and Frugal AI metrics must form part of a continuous optimisation cycle to manage an AI portfolio. This requires not only tracking costs but also measuring the value delivered, ensuring that every investment advances, both operational efficiency and strategic objectives.

An effective mechanism to accelerate responsible and efficient AI in the early stages of design is through sandbox environments, controlled low-cost spaces for experimentation, and rapid prototyping. UNICC's AI Sandbox provides such a platform, enabling users to test solutions through hackathons, agile prototyping, and collaborative experimentation. These environments can significantly reduce design and development costs, foster cross-user engagement, and accelerate progress toward SDG-aligned outcomes.

However, sandboxing must be integrated into a broader Design-to-Production lifecycle. This includes:

- **Clear criteria for progression** from prototype to production, based on ROI, adoption potential, and SDG contribution.
- **AI Governance checkpoints**[8] to ensure that sandbox learning is documented, reviewed, shared, and incorporated into future AI initiatives.
- **Scaling protocols** that allow successful prototypes to be industrialised quickly and costeffectively across Users.

In parallel with TCO measurement, this paper suggests that organisations should formally evaluate Return on Investment (ROI) for all AI initiatives, using the methodology outlined in Equation 4.2. This ensures that optimisation decisions are not purely cost-driven but balance expenditure with the tangible value created. ROI should be assessed at multiple stages of the lifecycle - post-pilot, post-implementation, and annually - to capture the evolving value profile of each solution.

---

[8] The mention of AI Governance checkpoints above is backed by Gartner's Market Guide for AI Trust, Risk and Security Management (AI TRiSM) (see [8]), which presents a framework to facilitate the safe, ethical, and compliant deployment of AI technologies. This framework consists of four interrelated layers, AI Governance, AI Runtime Inspection & Enforcement, Information Governance, and Infrastructure. It emphasises that successful AI risk management necessitates collaboration across various functions and investments in specialised tools.

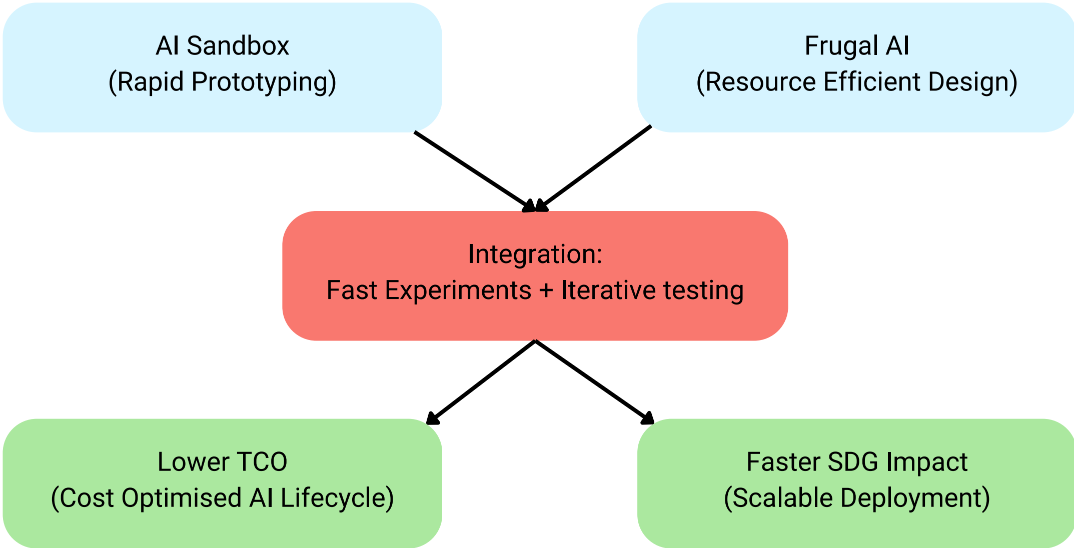


Figure 21: AI Sandbox + Frugal AI → Lower TCO and Accelerated SDG Impact.

Finally, sandbox initiatives should be explicitly linked to specific Sustainable Development Goals (SDGs), with measurable indicators to track contribution. This not only strengthens alignment with the UN's mission but also ensures that resources are directed toward projects with the highest strategic and social return. By embedding sandboxing within a structured measurement, feedback, and scaling framework, and by evaluating both TCO and ROI in line with SDG impact, organisations can position itself as a pioneer in resource-efficient and impact-driven AI innovation.

6.5 Related Work

The comprehensive measurement of AI portfolio value is a subject of growing importance, as organisations seek to move beyond fragmented assessments to a more holistic understanding of both costs and impacts. While numerous frameworks have been proposed in both academic and industry contexts, they often focus on specific dimensions of value. Our framework, as outlined in this paper, provides a unique, three-level approach that integrates TCO, ROI, and social impact aligned with the UNSDGs. This section situates our work by examining key contributions from related literature and highlighting how our model extends and complements existing approaches.

A useful point of comparison is the McIntyre and Liew Model [12], which presents a framework for evaluating AI initiatives in US federal agencies. Their model aims to quantify the value of an AI portfolio by transforming textual data from 1,754 publicly disclosed AI use cases into quantitative scores for benefits and costs, producing a relative ranking of projects based on ROI. While both papers share the objective of creating measurable ROI for AI projects, the distinction lies in scope and methodology. The McIntyre and Liew model provides a scalable, relative ranking tool for a large dataset, enabling broad comparisons across different government agencies. In contrast, our framework, developed in partnership with the UNICC, is a more prescriptive, absolute measurement tool tailored to a specific organisation's internal use. Crucially, our model goes beyond a purely financial ROI by introducing a third level of metrics that explicitly measures social impact and sustainability against the SDGs, a dimension not covered by the McIntyre and Liew model. This addition allows organisations to align AI value not just with financial returns but also with broader societal and environmental objectives.

The relevance of our approach is further reinforced by the MIT "State of AI in Business 2025" report [3], which provides a comprehensive analysis of AI implementation and introduces the concept of the "GenAI Divide." This divide highlights a stark reality: despite significant investment, 95% of AI pilots fail to deliver a measurable return. The MIT report identifies the core barriers as a learning gap (AI systems not improving with feedback) and a tendency to focus on visible but low-ROI front-office functions rather than high-impact, back-office automation. It concludes that successful implementations are led by external partnerships and are deeply integrated into specific workflows. These findings directly reinforce the strategic imperative of our framework. While the MIT report diagnoses why AI pilots often stall—lack of learning, poor workflow fit, and misguided investment—our paper provides a concrete framework for how to overcome these challenges. The proposed TCO, ROI, and SDG metrics supply the structured approach and quantitative grounding needed to bridge the GenAI Divide. By integrating our model, organisations can move beyond simply "investigating" AI to a measured, transparent process that aligns with both the business outcomes and social objectives highlighted in the MIT report.

6.6 Future work

The framework presented in this paper offers a comprehensive, three-level approach to measuring AI portfolios, integrating Total Cost of Ownership (TCO), Return on Investment (ROI), and alignment with the Sustainable Development Goals (SDGs). The following areas represent key opportunities for further work.

6.6.1 Expanding the Data and Empirical Modeling

As noted in the relevant areas of section 2, the empirical models in this paper are based on simulated datasets and a very limited number of real-world use cases, that are also specific to one organisation. To validate and enhance the predictive power of the models, it is essential to collect more datapoints either by statistically scaling the simulated datapoints by relevant techniques or by collecting datapoints from diverse sources such as other organisations. In particular, future research should focus on gathering a larger, more varied dataset that includes projects from different industries, geographic regions, and organisational structures. This would allow for the development of more robust and statistically significant regression models that can account for a wider range of variables and potential outliers, moving beyond simplified approximations to more accurately represent the complexities of real-world AI portfolios. Additionally, more datapoints could also mean different variables are more "more" significant than others in terms of p-value comparisons so the models may change the way they look. Of course, all of this is dependent on the new datapoints (especially those from the other diverse sources) not aligning with results presented in this paper.

Another natural step is to consider building on the introductory regression models. This could be done in a multitude of ways such as by either considering higher order terms for the cost modelling or exploring more factors / variables which could impact the score. Additionally, advanced AI-assisted methodologies for project cost estimation such as using deep learning models, as highlighted in Section 2.2.3 could be considered to help enhance forecasting accuracy, providing a more granular understanding of non-linear patterns and interactions between costs, value, and operational variables.

6.6.2 Relevance to Easily Accessible Data

The authors note that **tokens/joule** as a metric for measurement of Frugal AI is technically rigorous but not easily accessible for most organisations. In that light, simpler proxies such as compute hours, cloud energy dashboards, and emissions per API call can be used as practical alternatives for adoption of the suggestions in this paper. Note that the foundational principles remain the same regardless of the unit of measurement of the energy-based metric.

6.6.3 Deepening the AI Value Chain Analysis

This paper provides a high-level overview of the AI value chain. Further work could involve a more granular analysis of costs at different stages of the AI lifecycle and develop predictive models to optimise resource allocation in each stage. For example, researchers could explore how early-stage decisions in the design phase (e.g., choice of model architecture or data pipeline) have a compounding effect on TCO in the scaling phase. This would provide a more dynamic and actionable framework for managing an AI portfolio as a living asset, rather than as a static collection of projects.

6.6.4 Applying the Efficient Frontier to AI Portfolios

Viewing an AI portfolio as a strategic asset class, much like a financial portfolio, opens a new avenue for optimisation. In this context, the concept of the **efficient frontier** can be leveraged as a powerful analytical tool. Future work should explore methodologies to plot each AI use case on a scatter plot, analysing the interaction of **Level 1, Level 2 & Level 3** metrics. This visualisation would allow organisations to identify and aim at specific targets to achieve at each level with the portfolio.

By identifying projects that fall below the frontier, resources can be strategically reallocated, and investment decisions can be made to push the entire frontier upward over time. This approach would shift portfolio management from a reactive, project-by-project basis to a proactive, datadriven strategy aimed at maximising the overall value of the organisation's AI assets. By pursuing these avenues, the Frugal AI framework can evolve into a more comprehensive and powerful tool, enabling organisations to not only measure but also actively manage their AI portfolios for both economic efficiency and societal good.

As highlighted by Tomlinson et al. (2024) [18], large efficiency gains do not negate broader rebound or social impacts. Future extensions of this framework should therefore consider how reductions in per-task emissions interact with systemic impacts, ensuring that TCO and SDG metrics capture both direct efficiency and broader sustainability outcomes.

Survey Insight: Respondents noted that elevated coordination and documentation costs partly reflect the early maturity of the UNICC AI Hub. As governance processes stabilise and reusable templates are developed, these costs may normalise over time, suggesting a trajectory of efficiency improvements for future projects.

6.6.5 Future-Proofing the Framework

As AI paradigms evolve, the framework must be able to adapt to remain relevant and effective. In particular, three areas are likely to become increasingly important:

- **Agent-based AI:** the rise of autonomous and agentic AI systems requires cost and value models that capture the contribution of multiple agents operating across workflows. This shift calls for dynamic measurement approaches that reflect collaboration and orchestration, not just isolated project outputs.
- **Synthetic Data Costs:** the generation, storage, and validation of synthetic datasets is emerging as a major component of modern AI development. These costs must be integrated into TCO calculations, both as a financial expense and as an ethical consideration (e.g., bias propagation or false confidence).
- **Energy and Task-level Accounting:** while Section 4.6.4 introduces Task ROI, futureproofing requires extending this logic to energy use and emissions at the micro-task level. This is central to Frugal AI principles, where efficiency per unit of value becomes the key benchmark.

Incorporating these dimensions ensures the framework remains resilient and forward-looking. By anticipating shifts in AI practice, UNICC and its partners can continue to measure and manage AI portfolios in ways that align financial returns with sustainable development outcomes.

Conclusion

The authors propose in this paper a comprehensive set of metrics to measure AI portfolios as strategic organisational assets. By integrating **Level 1: Total Cost of Ownership (TCO), Level 2: Return on Investment (ROI),** and **Level 3: SDG Impact,** the framework provides a clear flow from costs to value to societal outcomes. This makes it accessible not only to technical experts but also to non-technical decision-makers across the UN system and beyond.

Beyond direct financial and operational considerations, the framework explicitly incorporates ethical costs and governance overheads. These are often hidden, but including them ensures a more credible and trustworthy basis for adoption.

The framework is designed to be **future-proof.** As AI paradigms evolve—such as agent-based systems, task-level ROI measurement, and the increasing role of synthetic data—the approach can adapt to remain relevant and effective. This flexibility allows organisations to continue measuring and managing AI portfolios as living assets, aligning financial returns with broader sustainability and inclusion objectives.

This initial white paper is part of a research program to gain further insights in implementation and change management.

Transparency fosters responsible use, and responsible use allows resources to be focused on AI initiatives that deliver the greatest possible value. Measurement should lead to optimisation processes that maximise the value of the AI portfolio over time, while also demonstrating alignment with the UN Sustainable Development Goals.

Our general approach is designed to be extended to other organisations. At the Frugal AI Hub, we will continue to gather feedback from organisations interested in piloting the framework, with the goal of refining metrics, sharing playbooks for adoption, and building an ecosystem of practice around Frugal AI.

References

[1] S. Abnar et al. Parameters vs FLOPs: Scaling Laws for Optimal Sparsity for Mixture-ofExperts Language Models. 2025. arXiv: 2501.12370 [cs.LG]. url: https://arxiv.org/ abs/2501.12370.

[2] L. Arga et al. “Frugal AI: Introduction, Concepts, Development”. In: Orange Research 27.1 (2025), pp. 72–111.

[3] A. Challapally et al. State of AI in Business 2025 Report. https://mlq.ai/media/ quarterly_decks/v0.1_State_of_AI_in_Business_2025_Report.pdf. Accessed: August 2025. 2025.

[4] B. Courty et al. mlco2/codecarbon: v2.8.2. Version v2.8.2. Dec. 2024. doi: 10.5281/zenodo. 14518377.

[5] Digital Public Goods Alliance (DPGA). Digital Public Goods Alliance. Accessed: August 17, 2025. 2025. url: https://www.digitalpublicgoods.net/.

[6] L. M. Ellram. “Total Cost of Ownership: An Analysis Approach for Purchasing”. In: International Journal of Physical Distribution & Logistics Management 25.8 (1995), pp. 4– 23.

[7] Forrester Research. AI Cost Optimization: The Why, What, And How. 2023.

[8] Gartner Research. Market Guide for AI Trust, Risk and Security Management. 2023.

[9] H. M. Gharaibeh. “Developing a Scoring Model to Evaluate Project Success”. In: International Journal of Computer Science and Information Security (2014).

[10] IBM. Smart Approaches to Allocations in Enterprise Financial and Operational Planning. 2023.

[11] International Telecommunication Union (ITU). AI for Good. Accessed: August 17, 2025. 2025. url: https://aiforgood.itu.int/.

[12] J. Liew and S. McIntyre. Introducing an AI-assisted Cost-Benefit-ROI Model for U.S. AI Use Cases. SSRN. 2024.

[13] NIST. NIST AI RMF Playbook. Tech. rep. Accessed: 2025. National Institute of Standards and Technology, 2023.

[14] Organisation for Economic Co-operation and Development (OECD). OECD.AI: Artificial Intelligence. Accessed: August 17, 2025. 2025. url: https://oecd.ai/en/.

[15] M. M. I. Shamim et al. “Advancement of Artificial Intelligence in Cost Estimation for Project Management Success: A Systematic Review of Machine Learning, Deep Learning, Regression, and Hybrid Models”. In: Modelling 6.2 (2025), p. 35.

[16] Swiss Sustainable Finance (SSF). Swiss Sustainable Investment Market Study 2025. Switzerland, 2025.

[17] N. Thompson et al. “A Model for Estimating the Economic Costs of Computer Vision Systems That Use Deep Learning”. In: Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 38. 21. 2024, pp. 23012–23018.

Appendix A. Other AI Providers: Cost Benchmarking

This appendix provides an indicative comparison of cost components across selected AI providers, using publicly available pricing data as of July 2025. Prices are shown in USD per one million tokens for both input and output. For consistency with the TCO framework, values may be converted to GBP using an exchange rate assumption of 1 USD = 0.78 GBP (July 2025).

Provider	Pricing Model	Representative Pricing (July 2025)		
Open AI	Primarily usage-based pricing, charged per million tokens processed, but also offers subscription plans (e.g., ChatGPT).	**Model**	**Input (USD/1M)**	**Output (USD/1M)**
		GPT-4.5 Preview	75.00	150.00
		GPT-4o	2.50	10.00
		GPT-40-Mini	0.15	0.60
Google Cloud	Mostly usage-based, pay-as-you-go pricing across text, image, audio, and video; subscription plans available for specific tiers.	**Model**	**Input (USD/1M)**	**Output (USD/1M)**
		Gemini 1.5 Pro	5.00	15.00
		Gemini Flash	3.50	10.50
Anthropic (Claude)	Tiered, usage-based pricing model; lighter models priced lower to encourage efficiency, with some subscription options available.	**Model**	**Input (USD/1M)**	**Output (USD/1M)**
		Claude 3 Opus	15.00	75.00
		Claude 3 Sonnet	3.00	15.00
		Claude 3 Haiku	0.25	1.25

Table 7: Representative pricing for OpenAI, Google Gemini, and Claude models (July 2025).

Note. *All figures are based on official pricing documentation from the respective providers as of July 2025.*

Appendix B. Dataset for Section 2.2

This table presents the datapoints used for modelling AI project value, including input cost factors, projected impact, and output metrics.

ID	Acquisition	Operational	Maintenance	Type	Months	Cost Score	Predicted
1	13.37	41.84	44.78	A	10	73.17462	53.85316
2	40.7	12.71	46.59	A	13	65.28845	88.14429
3	35.82	3.68	60.5	B	17	61.69905	62.8383
4	44.01	16.34	39.65	B	18	95.82473	97.11983
5	48.99	49.72	1.28	A	23	61.02158	78.62803
6	3.22	62.96	33.82	B	6	-34.17194	-19.01273
7	26.68	35	38.32	A	11	60.65835	62.63374
8	51.01	36.61	12.37	A	20	72.12379	83.33591
9	29.59	53.34	17.07	A	14	82.34883	53.09714
10	33.98	48.79	17.23	A	20	90.12768	72.09046
11	52.92	39.19	7.9	B	21	44.5554	39.91452
12	32.11	38.53	29.36	B	11	0.40742	15.86984
13	40.2	35.25	24.55	A	16	62.49022	74.72995
14	46.53	23.5	29.97	B	13	61.79744	34.67743
15	25.57	36.55	37.88	A	12	42.78084	63.66808
16	74.1	14.96	10.94	A	12	109.59808	83.60152
17	81.81	9.02	9.17	A	6	69.75827	74.29333

Table 8: Raw dataset of 17 projects underlying Section 2.2 cost modelling. Values shown to five decimal places for Cost Score and Predicted.

Appendix C. Dataset for Section 3.3

This table presents the datapoints used for the computation of the Net SDG Score.

Project	ROI	Reach	SDG Aspects	Energy Efficiency	Inclusion	Resilienc e	Harm Index	Net SDG Score
Project 1	3.33	96.67	22.49	69.86	29.81	70	44.52	42.14
Project 2	0.62	91.21	30.97	18.54	44.18	70	45.72	33.59
Project 3	0.56	72.16	51.94	44.56	65.15	30	30.44	43.6
Project 4	0.65	81.59	33.94	34.69	41.66	70	20.53	38.55
Project 5	1.35	12.22	87.2	83.32	20	70	7.35	37.43
Project 6	7.79	53	14.12	50.37	31.93	30	46.76	22.1
Project 7	1.71	78.26	49.8	82.91	70.13	30	15.06	60.21
Project 8	1.04	29.48	81.9	83.12	47.59	70	3.04	54.47
Project 9	2.2	38.64	20.97	81.49	80.94	30	47.39	39.71
Project 10	1.83	30.85	60.49	49.58	19.26	30	36.03	34.5
Project 11	1.05	22.85	28.59	77.9	49.14	30	7.11	33.81
Project 12	1.61	47.31	21.48	66.63	98.65	70	27.46	37.99
Project 13	1.23	47.24	77.8	73.92	90.37	70	47.7	58.77
Project 14	0.88	43.2	90.55	10.06	89.78	30	29.27	43.54
Project 15	1.94	23.72	43.7	52.78	25.75	70	20.23	26.52

Table 9: Raw dataset of 15 projects underlying Section 3.3 SDG Score. Values shown rounded off to 2 decimal places.

Appendix D. Selected Qualitative Insights

Theme	Representative Comments (anonymised)
Coordination & Documentation	“Significant effort needed for approvals and documentation across units.”
People Costs	“Senior architect and SME time dominated the design phase.”
Architecture Complexity	“Multi-tenant and multilingual support increased integration costs.”
Run Costs	“Run costs started low but grew steadily with increased usage and monitoring.”
Governance	“Compliance reviews and approval bottlenecks were a major overhead.”

Table 10: Illustrative qualitative survey insights on cost drivers.

Frugal AI Hub at Cambridge Judge Business School

Trumpington Street, Cambridge
CB2 1AG
United Kingdom

https://frugalai.org/

United Nations International Computing Centre

Palais des Nations
1211 Geneva 10
Switzerland

www.unicc.org