EGDC Case study – GlobalLogic Agentic Operation Management

October 2026

Case Study Methodology

EUROPEAN GREEN DIGITAL COALITION

Funded by the European Union

1 Introduction

The European Green Digital Coalition (EGDC) is an initiative of companies, supported by the European Commission and the European Parliament, based on the request of the EU Council, which aims to harness the enabling emission-reducing potential of digital solutions on all other sectors.

The main aim of the EGDC is to maximise the sustainability benefits of digitalisation within the ICT sector, while supporting sustainability goals of other key sectors such as energy, transport, agriculture, and construction. The Coalition recognises the need for science–based methods to estimate the reduction and avoidance of greenhouse gas (GHG) emissions by specific ICT solutions across sectors. This will accelerate the sustainability and circular transitions of these sectors while contributing to an innovative, inclusive, and resilient society.

To support the EGDC, a set of case study calculators are developed to provide a practical example of calculating the net carbon impact of a green digital solution in line with the European Green Digital Coalition (EGDC) methodology. This work aims to support the members of the EGDC with Action 2 of the EGDC Declaration.

This case study methodology accompanies the GlobalLogic Agentic Operation Management case study calculator and provides further details, additional context and transparency around the case study calculator to ensure the outcomes of the case study are interpreted and used correctly.

Funded by
the European Union

Disclaimer for European Parliament Pilot Project – European Green Digital Coalition (EGDC) Case Studies

The following disclaimer is intended to provide clarity and context for the case studies prepared as part of the EP Pilot Project, which have showcased the net carbon impact of specific digital solutions using the EGDC ICT Methodology developed during the project:

1. Purpose of the Case Studies:
The case studies served multiple purposes, including:

- **Development of the Methodology:** They contributed to the development of the EGDC ICT Methodology. These case studies were conducted concurrently with the methodology's creation and served as a valuable testing ground for its initial formulation.
- **Application Examples:** They provided practical examples of how the methodology can be applied to real-life use cases. These case studies were essential in demonstrating the practicality and effectiveness of the methodology when applied to concrete situations.
- **Identification of Improvement Areas:** By conducting these case studies, we aimed to highlight parts of the calculation in need of improvement. They shed light on the challenges and limitations inherent in using available data and indicated the necessary steps to move towards best practices in assessing net carbon impacts.

2. Data Quality as a Key Determinant:
It is imperative to emphasize that data quality is a fundamental determinant of the quality and reliability of the case studies. The accuracy and completeness of the data used significantly influence the outcomes and findings of these case studies.
It is essential to acknowledge that the data available for each case study may differ in terms of accuracy, granularity, and coverage. As a result, the case studies may not necessarily represent the best practice application of the EGDC ICT Methodology. Instead, they reflect the application of the methodology at various stages of data availability.

3. Liability for Errors/Omissions:
While reasonable steps have been taken to ensure that the information contained within the case studies is correct, the EGDC gives no warranty and makes no representation as to its accuracy. We accept no liability for any errors or omissions that may be present in the case studies, methodology, or related information. Users and readers are advised to exercise their judgment and seek further clarification if needed, as the information provided may evolve over time and depend on external factors beyond our control.

4. Appropriate Use of the Case Study Calculators:
The case study calculators are intended for educational and informational purposes. They rely on certain assumptions and input data to generate results.
The results of the calculators are specific to the implementation of the ICT solution and may not be representative for other implementation contexts.
As such, it is imperative for users to refrain from directly extrapolating these results to ICT solutions or implementation contexts that may seem conceptually similar.
Instead, users are advised to use the calculators as a means to understand the practical application of the EGDC ICT Methodology, thereby equipping themselves with the knowledge required to develop customized calculators specifically tailored to their unique ICT solutions and implementation circumstances.

In conclusion, these case studies provide valuable insights into the calculation of the net carbon impact of digital solutions through the practical application of the EGDC ICT Methodology. However, it is vital to exercise caution when interpreting the results, considering the variances in data quality and the evolving nature of the methodology. The findings are indicative of the methodology's potential and its room for refinement as we work towards more accurate and comprehensive assessments of net carbon impacts.

2 Results

ICT Solution and assessment overview	
GlobalLogic is a digital engineering services company and part of the Hitachi Group. It helps enterprises design, build, and operate digital products, software platforms, data systems, cloud solutions, and AI capabilities. GlobalLogic have developed the solution Agentic Operation Management (AOM) which is essentially an enterprise operating layer for AI agents. Rather than being an agent itself, it acts as the governance, orchestration, and management platform that sits above individual agents. The solution has been implemented within GlobalLogic's internal operations from December 2025 to April 2026. This case study is an ex-post assessment assessing the effect of AOM's model right-sizing (i.e. the practice of automatically selecting the smallest and least resource-intensive AI model capable of completing a task to the required quality standard, rather than always using the largest model).	
Quantified impacts	
Assessment period:	4 months, extrapolated to 1 year
Net carbon impact range:	USA: -13.7 to -3.8 kgCO2e/year Germany: +0.3 to +0.9 kgCO2e/year Combined: -3.6 to –12.8 kgCO2e/year

Organisational contribution
GlobalLogic is the developer of the solution and in the implementation context of this study is also responsible for deploying the solution. This aligns with A-level classification as defined by ITU-T L.1480 (contribution of implementing the integrated solution or the innovation of the solution).
Other identified impacts
Reduction in electricity use leads to significant cost savings. If all cost savings are invested into typical business carbon-intensive activities, it is estimated that this increase in emissions could outweigh the net saving. However, if GlobalLogic were to invest cost savings into the further development of energy optimisation solutions, rather than industry-average digital engineering services, emissions increases could be partially or completely negated. AOM can also reduce electricity consumption by optimising other processing tasks, for example by: caching existing answers, avoiding duplicate agents and duplicate computation, improving infrastructure utilisation, reducing failed or unnecessarily long AI runs. However, as the available data lends itself to evaluation of right-sizing, this is the sole focus of this case study.

Funded by the European Union

Net carbon impact:	USA: -7.3 kgCO2e/year Germany: +0.5 kgCO2e/year Combined: -6.8 kgCO2e/year	An additional longer-term impact may be a reduction in need for as much hardware if processing reductions are sustained.
Net impact per API request:	USA: -8.8×10^{-6} kgCO2e/API request/year Germany: $+0.6 \times 10^{-6}$ kgCO2e/API request/year Combined: -8.2×10^{-6} kgCO2e/API request/year	

3 Methodology

Name of solution	
Assessment Objective	The purpose of this assessment is to quantify the avoided emissions from AOM within GlobalLogic's operations by comparing energy consumption per API request in the implemented AOM scenario against a counterfactual scenario without AOM. The output of the assessment will demonstrate the solution's impact on energy consumption and CO_2 emissions reductions. The assessment will focus on a single implementation context: Agent usage within GlobalLogic's own operations over a fourth-month time period, with processing occurring in Google cloud regions of Iowa, USA or Frankfurt, Germany. An ex-post assessment will be conducted, drawing on operational data from before and after solution deployment to establish a robust comparison. The assessment will be extrapolated to an annual impact.

Funded by the European Union

Solution Description	Agentic Operation Management (AOM) is a platform that provides a shared operational layer for AI agents across different frameworks, models, and deployment environments. It enables organisations to create, deploy, connect, monitor, and govern AI agents through a common control plane, including agent cataloguing, secure communication, model routing, monitoring, and policy enforcement capabilities. The focus of this case study is on the net carbon impact of model right-sizing, whereby each task is routed to the smallest suitable AI model rather than defaulting to more resource-intensive models. The magnitude of emissions reductions depends on the type of AI workloads being managed and how the user makes use of the solution. The solution provides information on the performance and token intensity of routing requests to different sized models, however, it is ultimately up to the user on whether to adhere to routing to the recommended model size or not.
Solution Boundary	The solution comprises of core digital components that are shared system-wide infrastructure: • Central control plane • Background scheduler • Admin UI • LiteLLM gateway These components orchestrate and route traffic for all agents globally and they are not mapped to any specific agent.
Functional Unit	Per API request was selected as the functional unit because it represents the same user outcome in both the reference and solution scenarios. This enables a like-for-like comparison of the energy consumption and associated emissions of processing an equivalent task, while isolating the

Funded by the European Union

	impact of model right-sizing from changes in activity levels or user demand.
Calculation Boundary	The calculation represents a one-year period by extrapolating the results from a four-month dataset, assuming that the GlobalLogic operations and usage of the solution would remain consistent annually. The API requests that were conducted within the dataset have been mapped to Google Cloud regions of either us-central1 (Iowa, USA) or europe-west3 (Frankfurt, Germany). It is therefore assumed that the electricity generation emission factors published for each Google Cloud region are the most appropriate to use to calculate the emissions impact of electricity consumption associated with token processing.
Reference scenario	The reference scenario comprises of a counterfactual situation where GlobalLogic employees manually chose which AI model to send tasks to based on judgement of what may produce a suitable balance of quality output and cost of processed tokens. The same API requests would be conducted as per the solution scenario, however, AOM would not be available to provide routing recommendations based on assessment of to the smallest suitable AI model. In theory, the reference scenario leads to a greater proportion of requests being sent to models that are more resource-intensive than necessary to produce the desired output. Simple API requests that only require a small number of tokens when right-sized to a light model can consume the same number of tokens as a complex API request when sent to a resource-intensive model. In this case study, data was unavailable relating to the proportion of API requests that were sent to each model size prior to the implementation of AOM. API requests and the models they were sent to have only been tracked since the introduction of the solution. To form a basis for the reference

Funded by the European Union

	scenario, the case study defines the starting point of likely proportion of API requests being sent to low (equivalent to Gemini Flash-Lite models), medium (equivalent to Gemini Flash models) and high (equivalent to Gemini Pro models) resource-intensity models. For GlobalLogic, the distribution has been assumed as 15% (low), 50% (medium) and 35% (high) in this case study. This is because GlobalLogic's core business is engineering services, digital product development, embedded software, data/AI solutions, and complex enterprise transformation. For these activities, medium-size models would likely be the workhorse model for most engineers. However, usage of high-size models is also significant as GlobalLogic's activities likely generate a larger share of work that benefits from advanced reasoning, code generation, architecture analysis, and agentic workflows than a typical IT services firm. Therefore, the proportional split of 15:50:35 is taken as a reasonable representation of a market average reference scenario. However, it is the greatest source of uncertainty within the model, and it is recommended to view results with a range of different reference scenarios based on this input to view the possible variability in results. This has been reflected within the uncertainty analysis of this case study.
Description of 1st order effects	The solution comprises of core digital components that are shared system-wide infrastructure: • Central control plane • Background scheduler • Admin UI • LiteLLM gateway The electricity consumption associated with orchestrating and routing traffic for all agents has been recorded in the dataset. The emissions are calculated based on which Google Cloud region the electricity consumption occurred in. It is assumed that no additional physical infrastructure was needed to deploy the AOM

Funded by the European Union

	solution, and therefore embodied emissions of any hardware that supports the running of the solution is not included in the first order effects.
Description of 2nd order effects	AOM is expected to reduce computational processing and associated electricity consumption (and hence emissions associated with electricity generation) through several mechanisms: 1) Model right-sizing, whereby each task is routed to the smallest suitable AI model rather than defaulting to more resource-intensive models. 2) Answer caching and reuse, which reduces duplicate computation by serving previously generated responses where appropriate. 3) Consolidation of AI workloads onto shared infrastructure, increasing utilisation and reducing redundant resource provisioning. 4) Reduction of failed or unnecessarily long AI runs through controls such as pre-checks, circuit breakers, and timeouts. 5) Reuse of existing agents through a shared catalogue, reducing duplication of AI services and associated compute requirements.
Description of higher order effects	Right-sizing models should lead to a reduction in the average number of tokens that are produced and processed per API request. Tokens have a unit cost and are generally more expensive when generated as part of a request to a more resource-intensive model. Therefore, reducing the volume of tokens and routing to lighter model sizes can significantly reduce costs of using AI model services. Cost savings could be invested into a variety of different activities which could potentially act counter to the net carbon savings. Over a long-term timescale, production and waste emissions may be avoided by virtue of optimising energy demand such that fewer pieces of hardware are needed to perform the same functional output and that in operation is operated with smoother, right-sized profiles, reducing thermal stress and extending lifespan. For this impact to materialise in a noticeable manner, it would likely take

	widespread sustained adoption of the solution across multiple companies.

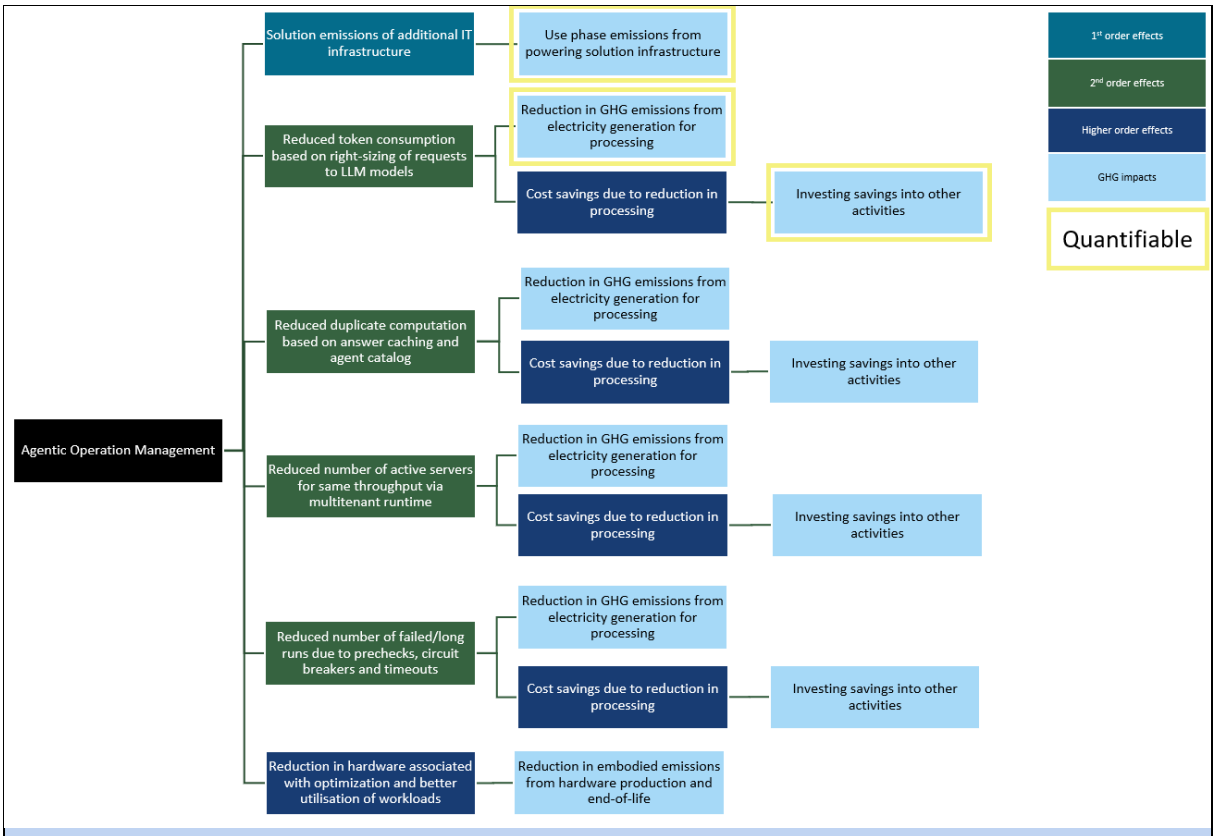
Assessing the impact of higher order effects

The below table shows the potential carbon impact of investing all the net annual cost savings into typical industry activity rather than ringfencing the financial savings for carbon saving activities.

Higher Order Effect	Quantitative Assessment	Magnitude	Likelihood	Mitigation
Cost savings due to reduced energy consumption	**Annual net cost saving (USA):** $4,167.70/year **Annual net cost saving (Germany):** $31.54/year **Total ~~stimated~~estimated emissions impact:** 425.8 kgCO2e/year (Calculated using: Open CEDA by Watershed 2025 emission factor for computer systems design services)	**High:** Two orders of magnitude greater than the net carbon savings from the solution	**High:** unless cost savings are ring fenced for specific activities	If GlobalLogic were to invest cost savings into the further development of energy optimisation solutions, emissions increases could be partially or completely negated.

Mapping of all effects

Funded by the European Union



Description of calculation	**First order effects** The solution emissions are calculated by summing the energy consumption recorded against all rows in the source dataset relating to system-wide added infrastructure that supports with the routing of the API requests. There are rows in the dataset that relate to infrastructure energy consumption but only those which are additionally required to facilitate the solution are included in the first order effects. The total energy consumption from the solution is then divided by the total number of API requests extrapolated to quantify the first order effects on the basis of the functional unit. **Second order effects** The difference in emissions between the reference and solution scenarios is calculated by deriving a weighted average energy per unit of workload under each scenario. To do this, the four-month dataset record of all API requests from GlobalLogic's operations is used to identify the total number of tokens and API

Funded by the European Union

	requests (and hence average tokens per API request) that were handled by each AI model type. From the allocation split to each AI model type, a weighted average of tokens per API request is calculated, which is then multiplied by a weighted average of the energy consumption per token to produce a value for the total energy consumed per API request under the **solution** scenario. To generate an equivalent value of total energy consumed per API request under the **reference** scenario, the assumption is that the total number of API requests processed would remain the same but distribution of API requests per model type are different. As described in the 'Reference Scenario' section, it is assumed that prior to the introduction of AOM, API requests would be routed to low, medium and high resource intensity models in a ratio of 15:50:35, respectively. The different AI model types that were used by GlobalLogic during the assessment period are each mapped to either low, medium or high based on the capabilities of each model type, with the suites of Gemini Flash-Lite, Gemini Flash and Gemini Pro, representing the different source intensity gradings from low to high. Therefore, for the reference scenario, the total number of API requests are then redistributed across the model types based on the mapping and the relative usage of each model type within each grouping in the solution scenario. Another key working assumption behind the calculation of second order effects is that the average token consumption per API request (and hence energy consumption per API request) for each model type remains constant between reference and solution scenarios; it is just the distribution of API requests between model types that is assumed to change. The premise here is that simple requests that require a small number of tokens when routed to a low resource intensity model, can be treated in the same manner as a complex request when sent to a high resource-intensity model and therefore, generate just as many tokens as a complex request. Under this assumption, the reference scenario weighted

Funded by
the European Union

	average energy consumption per API request is calculated. As a final step, the second order effects are calculated by subtracting the weighted average energy consumption per API request in the solution scenario from that in the reference scenario. **Net Carbon Impact** The net carbon impact is calculated by subtracting the first-order emissions generated by the solution itself from the second-order effects. This ensures that the environmental benefit accounts for the emissions embedded in the solution components and cloud usage, providing a more accurate measure of the net reduction in greenhouse gas emissions. The calculations steps are repeated separately for the two different geographic regions and extrapolated such that the net carbon impact can be annualised. Multiplying the normalised net carbon impact (i.e. per API request) by the total number of API requests processed gives the estimated absolute net carbon impact over the course of a year. It should also be noted that although the combined results for the two geographic regions creates a net carbon impact reduction, the net carbon impact from energy change in Germany is actually an increase. This is likely reflecting the high dependency on users to respond to solution recommendations such that right-sizing optimisation is implemented in practice; there may be instances where this is not done and an energy/emissions reduction is not realised.
Net Carbon Saving Impact of the Solution	1st order effect (emissions from solution components): 0.1 kgCO2e/year 2nd order effect (avoided emissions from reduced energy use): - 6.9 kgCO2e/year Total net carbon saving impact: -6.8 kgCO2e/year **Other Savings from reference scenario:**

Funded by the European Union

	% Energy savings through use of the solution: 39.3% CO2 emissions reduction per API request per year: -8.2 x 10^{-6} kgCO2e/API request/year The results above relate to the combined impact across geographic regions. Please see the Results section for the regional breakdown.
Uncertainty and sensitivity analysis	**Total carbon savings enabled ($kgCO_2e$/year):** Calculated Net Avoided Emissions: -6.8 Lower uncertainty range: -3.6 Higher uncertainty range: -12.8 The results above relate to the combined impact across geographic regions. Please see the Results section for the regional breakdown. **Uncertainty & Sensitivity Analysis:** The key source of uncertainty within establishing the electricity consumption reduction is in defining how API requests would have been processed across different types of AI models prior in a counterfactual without AOM. Varying the reference scenario by adjusting the allocation of API requests to low, medium and high intensity AI models across different possibilities of decreasing likelihood indicates quite a broad range of results for the net carbon impact, however, most realistic inputs for a company like GlobalLogic could lead to a net carbon reduction per API request. The uncertainty analysis reflects this low level of reliability of the reference scenario activity data, and this contributes to the overall uncertainty range stated above. The factor which when changed has the biggest impact on the net carbon impact is the electricity reduction that occurs between the reference and solution scenarios. Electricity consumption in the US is the most material contributor and therefore

Funded by the European Union

	has the highest impact within the sensitivity analysis. Given the high level of uncertainty in the reference scenario data inputs for this parameter, it was determined sensible to evaluate the sensitivity response based on the uncertainty analysis output that US electricity consumption could vary between +50% to -33%. Inputting these upper and lower bounds into the sensitivity analysis, leads to a +54% to -36% variation in the net carbon impact from the reduction in electricity.
Assumptions	**Reference scenario model routing**: In the absence of pre-AOM data on how API requests were distributed across model sizes, the reference scenario assumes that requests would have been routed manually across low, medium, and high resource-intensity models in a ratio of 15% low, 50% medium, and 35% high. This is treated as a reasonable counterfactual for a digital engineering services company such as GlobalLogic but is recognised as a key source of uncertainty. **Functional unit**: The assessment uses per API request as the functional unit, on the basis that each API request represents the same user outcome in both the reference and solution scenarios. This enables a like-for-like comparison of the energy consumption and associated emissions of processing an equivalent task, while isolating the impact of model right-sizing from changes in activity levels or user demand. **Assessment period and annualisation:** The calculation is based on a four-month operational dataset and extrapolated to represent a one-year period. This assumes that API request volumes, model usage patterns, workload characteristics, and regional distribution of processing remain broadly consistent over a full year. **Calculation boundary**: The assessment focuses on API requests processed within GlobalLogic's own operations, with processing mapped to Google Cloud regions in Iowa, USA and Frankfurt, Germany. The results are therefore specific to this implementation context and should not be directly extrapolated to other organisations, workloads, cloud regions, or deployment configurations without adjustment.

Funded by the European Union

Model category mapping: AI models used in the assessment are mapped to low, medium, and high resource-intensity categories, with Gemini Flash-Lite, Gemini Flash, and Gemini Pro model families used as representative equivalents for these categories. This mapping is used to compare the model mix in the solution scenario with the assumed model mix in the reference scenario.

Constant workload demand: The calculation assumes that the total number of API requests is the same in the reference and solution scenarios. The assessed impact therefore reflects a change in the energy intensity of processing those requests, rather than a change in the level of demand or user activity.

Constant token intensity by model type: The average token consumption per API request for each model type is assumed to remain constant between the reference and solution scenarios. The difference between scenarios is therefore modelled through a change in the allocation of requests between model categories, not through a change in the token generation behaviour of each model type.

Solution scenario based on observed operation: The solution scenario uses observed API request and model usage data from GlobalLogic's operations following the implementation of AOM. This reflects the routing of requests under the implemented solution and is used to calculate the weighted average energy consumption per API request in the solution scenario.

First-order effects: The first-order emissions include the electricity consumption of additional system-wide infrastructure required to support AOM, including components such as the central control plane, background scheduler, admin UI, and LiteLLM gateway. Only infrastructure considered additional to the solution is included in the first-order effects for the calculation.

No additional embodied hardware emissions: It is assumed that no additional physical infrastructure

Funded by the European Union

	was required to deploy AOM. As a result, embodied emissions associated with hardware manufacture, transport, installation, maintenance, or end-of-life treatment are excluded from the first-order effects. **Regional emissions calculation**: Electricity-related emissions are calculated based on the Google Cloud region in which the electricity consumption occurred. The model therefore applies region-specific electricity emissions factors to the energy consumption associated with processing in Iowa, USA and Frankfurt, Germany. **Scope of quantified second-order effects**: The quantified second-order effect focuses on model right-sizing, whereby API requests are routed to the smallest suitable AI model rather than defaulting to more resource-intensive models. Other potential optimisation mechanisms, such as answer caching, duplicate computation avoidance, improved infrastructure utilisation, failed-run reduction, and agent reuse are identified but not quantified in the main result due to the available data being most suited to evaluating model right-sizing. **Cost saving reinvestment**: The assessment recognises that reduced energy consumption can create cost savings, which could lead to rebound effects if invested in carbon-intensive activities which could offset the emissions reduction. If savings are instead reinvested into additional energy-saving solutions, rather than into industry-average business activity, a net emissions reduction effect could be preserved. **Higher-order effects**: Potential longer-term impacts, such as reduced hardware requirements from sustained reductions in processing demand, are identified qualitatively but are not quantified in the main net carbon impact result. These effects would require longer-term evidence of sustained adoption and measurable changes in hardware procurement, utilisation, or replacement cycles.

Funded by
the European Union

Data sources	**Data provided by GlobalLogic:** • Four-month organisational record of AI model workloads following the introduction on AOM, including data on volume API requests made, volume of tokens processed, cost of tokens processed and energy consumption. **Other data sources:** • All secondary data sources used during calculations of the 1st order effects have been linked within the calculator
Input adjustments and key considerations for usage of results	The results are specific to GlobalLogic's AI workloads, model mix, cloud deployment regions and assumed reference scenario. When applying this methodology in other implementation contexts, users should recalculate inputs relating to baseline model allocation, API request volumes, token intensity, model portfolio, regional electricity emission factors and infrastructure configuration. Additional benefits from optimisation measures such as caching, duplicate computation avoidance and agent reuse should also be considered where relevant.
'Do no harm' criteria	Do not foresee any negative impacts on any of the EU Taxonomy's environmental nor social objectives and strongly supports objective 1: climate change mitigation.
Key areas for improvement	The largest opportunity for improvement is the collection of measured pre-implementation data on model-routing behaviour, reducing reliance on assumptions within the reference scenario. Future studies should also expand the assessment beyond model right-sizing to include other optimisation mechanisms enabled by AOM, such as answer caching, duplicate-computation avoidance, workload consolidation and agent reuse. Additional improvements could be achieved through longer monitoring periods, quantification of hardware lifecycle impacts, and measurement of higher-order effects such as reinvestment of cost savings.

Funded by the European Union