Why Measuring AI Environmental Impact of Organisations is Non-Trivial?

Loïc Guibert [1] **David Bekri** [1] **Louise Aubet** [2] **Steve Berberat** [3] **Sébastien Rumley** [1]

Abstract

This paper presents the results and conclusions from the *EIEIAE* project, which defined a methodology to measure the environmental impact of AI services within companies and organisations. Several lessons can be learned from this project, as it highlighted several challenges that emerged when trying to produce reliable estimates of such impacts: (a) the industrial lack of transparency of AI providers, (b) the absence of accurate and exhaustive modelling of ICT environmental impacts, (c) the inherent complexity of gathering the necessary data for organisations, and (d) the fact that organisations are more focused on cost savings than on reducing their environmental impact.

1. Introduction

Information and Communication Technologies (ICT) represent a significant share of global resources usage. In 2023, the sector consumed an estimated $1,000\ TWh$ of electricity, accounting for approximately 4% of worldwide electricity use (International Energy Agency, 2025). Data centres alone accounted for $360\ TWh$, therefore representing 1.5% of such electricity use. Despite latest efficiency gains, the share of data centres electricity consumption is expected to raise up to 3% in 2030 (International Energy Agency, 2025), notably due to the recent surge of Artificial Intelligence (AI). Other studies project that the energy consumption of data centres will increase by at least 10% per year until 2030 (McKinsey, 2023), and could even rise from $530\ TWh$ in 2023 to $1,490\ TWh$ in 2030 (The Shift Project, 2025). Such figures underscore that accounting the environmental impact of ICT is a critical priority, particularly regarding the escalating resource demands of AI.

The pursuit of financial optimisation drives organisations to streamline operational costs, which includes ICT and AI infrastructures. This economic imperative often converges with corporate sustainability objectives, as lower energy requirements concurrently reduce environmental impact. Moreover, the regulatory landscape of sustainability has undergone a major shift with the arrival of the Corporate Sustainability Reporting Directive (CSRD) (Hummel & Jobst, 2024). This European directive establishes a mandatory disclosure regime that obliges many European companies and some of their suppliers to provide audited, standardised reports on their environmental and social impacts.

Nevertheless, quantifying the environmental impact of ICT remains a significant challenge for organisations, particularly due to the distributed and often opaque nature of remote and cloud-based services. Assessing the specific impact of AI introduces additional complexity. To date, there is no widely accepted framework that defines how to quantify the environmental impact of AI in practice within an organisation. The *EIEIAE* project addressed this gap in existing evaluative frameworks, and aimed to analyse how to concretely assess the environmental impact of AI, both remotely and on-premises, within organisations. Finished in 2025, its research question was: *How can we quickly and affordably analyse the environmental impact of AI in an organisation while maintaining the accuracy of the results within an acceptable range?*

The project was organised around four phases, each dedicated to: (1) assessing organisations' needs for quantifying the environmental impact of AI; (2) analysing the state of the art in methods for measuring the environmental impact of AI; (3) developing a methodology for the rapid assessment of AI-related environmental impact within an organisation; and (4) validating the methodology with a concrete use-case. This paper summarises the results of the project.

2. Organisations Needs

With the objective to accurately identify both the needs and opinions of organisations about AI environmental impact, semi-structured interviews have been conducted within 9

[1]iCoSys, HEIA-FR, HES-SO University of Applied Sciences and Arts Western Switzerland, Fribourg, Switzerland [2]Resilio SA, Lausanne, Switzerland [3]HEG Arc, HES-SO University of Applied Sciences and Arts Western Switzerland, Neuchâtel, Switzerland. Correspondence to: Loïc Guibert <firstname.lastname@hefr.ch>, David Bekri <firstname.lastname@hefr.ch>, Steve Berberat <firstname.lastname@he-arc.ch>, Louise Aubet <firstname.lastname@resilio-solutions.com>, Sébastien Rumley <firstname.lastname@hefr.ch>.

Proceedings of the Swiss AI Days 2026, Martigny, VS & Fribourg, FR, Switzerland PMLR 309, 2026. Copyright 2026 by the author(s).

companies. Interviews took place at the end of 2024 and at the beginning of 2025. Different profiles have been selected with the goal to dress an accurate variety of commercial services they offer, summarised in Table 1. We requested interviews with managers from the selected companies, or with employees involved in sustainability aspects.

It must be emphasised that there is only one company fitting the second category: the results of the survey relative to this category must therefore be interpreted with caution.

Table 1. Summary of the interviewed companies to determine their needs. The description explains which type of services are proposed by each category, alongside the amount of interviewees of the said category.

Cat.	Description	#
1	Services without any relation to AI	4
2	Services integrating existing AI	1
3	Developing their own AI services	4

To ensure consistency and internal validity, a standardised interview guide was employed across all sessions, defining a set of questions that were then asked. The questions can be consulted in Appendix A. However, the semi-structured approach allowed for asking questions and driving open-ended discussions when interviewees raised points of particular relevance to the research objectives. This approach ensured that while core thematic areas were systematically addressed, the unique insights of each interviewee were also captured. To bridge the gap between qualitative nuance and statistical clarity, notes were taken in a summarised format, whereby specific themes or responses were tracked for frequency and intensity alongside meaningful information.

The restricted sample size prevents formal statistical validation, thereby situating the findings as exploratory in nature rather than generalisable across a broader population.

2.1. Employees

Overall, 4 interviewed companies employ between 11 and 35 people; one company employs more than 2000 people; and 3 of them only employ 4 to 10 people. 7 of the companies are mainly active in ICT services, and 3 of them employ at least one data scientist.

2.2. Sustainability Policies

Only one of the surveyed companies, member of the third category, has a genuine sustainability policy in place. The other companies do not have any sustainability policy, but some have implemented environmental-related practices. One company was in the process of implementing such a policy.

Sustainability policies are implemented for internal reasons, thanks to management interest or employees commitment, or for external reasons, answering to customer requests or for brand image purposes. 3 companies cited both reasons for putting a sustainability policy in place. 2 companies cited employee satisfaction as the main driver, and 2 companies cited customer satisfaction. Companies that did not cite external reasons as decisive confirmed that their customers had never shown any explicit interest in such subject. It should be noted that one of the interviewees pointed out that a company cannot implement efficient sustainable actions if its management does not believe in the cause.

None of the surveyed companies had been asked to provide information on corporate sustainability in invitations to tender, except for one. However, according to the respondent, this was a case of greenwashing.

The concept of CSRD is little known among surveyed companies, regardless of their category. Only one company is compliant with the directive, and only one respondent had heard of it. Other respondents were unaware of its existence.

Only one company has ever calculated its carbon footprint, and another company has once calculated the carbon footprint of its cloud services; both of them are from category 3. No other companies have ever performed such calculations.

The needs of companies' clients on environmental reporting is quite low, with little differences among categories:

- **Category 1**: Negligible interest (1 request per 1,000–5,000 clients); One company did not respond.
- **Category 2**: Little interest.
- **Category 3**: Occasional to rare informal requests; Primary concern is data security over sustainability.

2.3. ICT and AI Costs

The proportion of ICT-related costs ranged from 2% to 35%, with respect to their turnover. Companies in categories 2 and 3 have higher proportions than those in category 1, thus implying that integrating AI into a company's services or products increases the proportion of ICT costs. Subscriptions to external AI services range from CHF 600 and CHF 1,000 per year, which is relatively low.

The proportion of overall ICT costs dedicated to AI also highly depends on the category:

- **Category 1**: either none, CHF 1,000 of costs, or not calculated.
- **Category 2**: 10% of costs.
- **Category 3**: either none, 30-40% or almost 100% of costs.

Category 3 costs certainly depend on the type of service offered by the company, as well as its engagement in research and development activities. One company is evaluating the possibility of purchasing GPUs, which could increase their AI-related costs.

2.4. Needs for Environmental Impact Audits

Three companies responded that they were not interested in performing audits on their environmental impact caused by AI, citing reasons such as not having a dedicated sustainability budget or not having fully completed the implementation of AI in their services or products.

Other companies responded that they would be interested in such a solution, but it would depend on the cost of the solution: this aspect was often mentioned as a strong hurdle. Two companies mentioned amounts that could be bearable for them: one mentioned a maximum price of CHF 3,000, while the other considered an amount between CHF 5,000 and CHF 10,000. These prices should be considered in relation to the companies' budget, which varies greatly.

One respondent indicated that their personal interest for such an audit is bigger than the one of their company. Another person felt that such an audit would be too expensive, and that their company's priorities lay in certifications related to data protection and security.

In contrast, a respondent considered that this audit would be useful in supporting internal decisions. Another respondent indicated that such audits are useful for measuring the impact of processes, with the aim of then implementing measures to reduce such impact. A third respondent expressed doubts that their company would be willing to fund such a service, particularly if it came at the expense of security-focused audits.

2.5. Miscellaneous

The interviews yielded additional relevant insights.

One respondent noted that their business model conflicts with sustainability targets because their primary revenue stream inherently relies on polluting services.

Another person indicated that one of their colleagues travels by plane within Europe about once a month, implying that reducing ICT impacts would certainly not be the most effective lever.

One respondent stated that it would rather be up to their suppliers to provide a balance sheet, rather than analysing their own emissions.

One person indicated that carbon emissions associated with their AI models are distributed equally between model training and inference. This parity suggests that both lifecycle phases can exert a comparable environmental influence depending on specific operational parameters.

Interestingly, several respondents admitted that the interview had increased their awareness of AI's monetary costs more than its environmental impact, suggesting that consulting services aimed at reducing AI costs may attract more clients than services solely focused on reducing environmental impact. In this respect, financial efficiency emerges as a potential lever for improving sustainability levels.

3. State of the Art

To define a methodology, the existing landscape of approaches assessing the impact of AI services has been explored. This Section discusses the most important findings we discovered during our gathering and analyses.

3.1. Scope

Figure 1 lists all the ICT subjects considered for the methodology scope, with included and excluded subjects. Subjects that can not be assessed in all contexts without establishing strong or even impossible hypotheses have been excluded.

Although it would be theoretically possible to calculate carbon emissions due to GPU manufacturing used by remote AI services, the lack of provider transparency on their infrastructure, amount and types of GPUs makes it impossible to precisely determine trustful figures. Likewise, emissions caused by the training phases of AI models are not disclosed, particularly for foundation models trained by big companies due to the lack of available data. This situation thus complicates the integration of training impact in the methodology. However, such calculation is possible for on-premises infrastructures, where data is accessible.

Furthermore, modern ICT services share their computing platforms with several other processes, a uncomfortable situation with opaque boundaries requiring complete and complex attributional Life Cycle Assessments (LCA) to be conducted.

Lots of tools, calculations and methodologies attempt to estimate energy consumption and carbon footprint of AI services, infrastructure and computation processes. However, as it is the case for ICT in general, the state of research for other types of impacts, such as water and abiotic depletion, is still relatively unexplored.

3.2. Existing Methods

Because of the aforementioned challenges, different methods have been defined by the research community, categorised based on their approach of estimating environmental impacts. The most adapted one are then selected for our methodology.

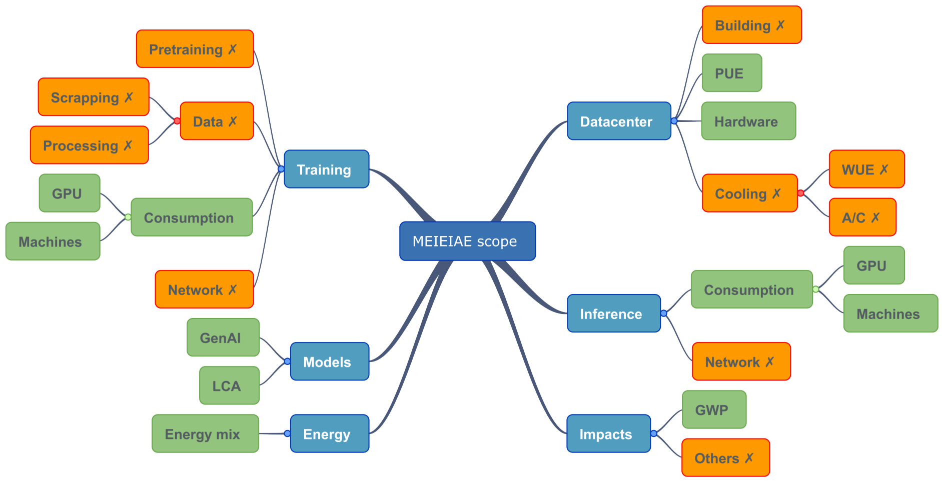

Figure 1. Scope of the project methodology. First-level nodes, in blue, represent the most impactive categories of ICT, with children and grandchildren leaves representing more specific subjects. Green leaves represent the subjects included in the scope, orange leaves represent the excluded subjects; the latter also feature crosses to ensure visual accessibility.

3.2.1. REQUEST-BASED

Several software libraries exist for measuring the environmental impact of processes running on computers and servers. Some of them include GPUs support, specifically allowing to measure the impact of AI services. The most popular library is CodeCarbon (CodeCarbon, 2025), but a multitude of similar alternatives exists (Boavizta, 2025). Other software libraries are designed to measure the environmental impact of remote AI requests through providers' API, such as the EcoLogits library (Ecologits, 2025).

Other tools exist as online calculators, with the most popular ones summarised in Table 2: (1) AI Emissions Scenario (Borisruf, 2025), (2) EcoLogits Calculator (Rincé et al., 2025), (3) AI Energy Score (Hugging Face, 2025), (4) LLM-Perf Leaderboard (Optimum, 2023), (5) ML.ENERGY Leaderboard (ML.ENERGY, 2025), and (6) LM CO_2 impact (Lacoste et al., 2019). EcoLogits being the most complete and precise in its calculation methodology, we selected it as the approach to be used for the requests-based method. More details about this tool can be consulted in Appendix B.

Green Algorithms calculator (Lannelongue et al., 2021; 2026) is another alternative complementary to the previous ones, based on request execution time and offering a more advanced and generic calculation method best suited for well-known infrastructures. Specific hardware can be specified, and the learning phase impacts can also be included.

Table 2. Characteristics of online calculators based on requests. HW = hardware type can be selected; PR = prompt size can be specified; RE = hosting region can be specified; IM = environmental impacts are provided alongside energy; PA = amount of parameters can be specified; EM = embodied carbon is considered.

Tool	HW	PR	RE	IM	PA	EM
(1)	✗	✓	✓	✗	✗	✗
(2)	✗	✓	✓	✓	✓	✓
(3)	✗	✗	✗	✗	✓	?
(4)	✓	✓	✗	✗	✓	?
(5)	✓	✗	✗	✗	✓	✗
(6)	✓	✗	✓	✗	✗	✗

3.2.2. COST-BASED

An assumption considers that the fees charged by AI providers to their customers cover only the inference part of their services, meaning that operational costs resulting from their infrastructure are offset by this revenue. In this case, these operational costs represent electricity costs, where large-scale fundraising and investments by major private groups are used to reinvest in the research and development phases for new models, while covering salaries, equipment, GPUs, and related costs (Zitron, 2025; Nover, 2024; Pineda et al., 2024).

Following this hypothesis, we assume that the combination of the flat rate limits and the price paid for subscriptions allow to approximate the upper threshold that can be assumed by AI providers. It serves as a strong heuristic and its sturdiness can not be verified with current existing knowledge.

3.2.3. Provider-based

ICT service providers often offer control panels to their users, mainly used for billing purposes and resources management. Such panels often allow access to monitor resources consumption.

Most control panels supported by providers adopt the Greenhouse Protocol (GHG) (Iwata & Okada, 2012; Safaei, 2011), which is an exhaustive standard with a ICT specialisation including GPUs (GeSI, 2017; Mytton, 2020), allowing environmental impact calculations to be used out-of-the-box (Google Cloud, 2026b; Microsoft, 2026; Amazon Web Services, Inc., 2026; OVH, 2023).

Other providers only provide information on utilisation statistics, such as the amount of requests counted by their systems. This is mainly the case for AI service providers: such figures can be combined with cost-based methods to calculate environmental impact.

3.2.4. On-premises

The on-premises context brings little challenges when it comes to assess environmental impact, regardless if it concerns ICT or AI services. The most recognised assessments have been selected based on the research question objective.

4. Methodology

Based on the methods presented in Section 3, we have established a methodology for evaluating the environmental impact of AI in organisations. Figure 2 shows its composition, covering as many situations as possible.

The three remote methods are based on costs, on the requests count, and on providers data. They fit in the scope defined in Figure 1. They can be used complementarily when services are used on different locations. A fourth method is dedicated to on-premises scopes, followed by some useful, generic calculations.

This methodology has been defined with the specific constraint to respect an effort-precision trade-off inherited by the research question, resulting of maximising the impact accuracy while being quickly calculated.

If no information are available when applying methods and equations, useful data drawn from scientific literature can be consulted in Appendix C.

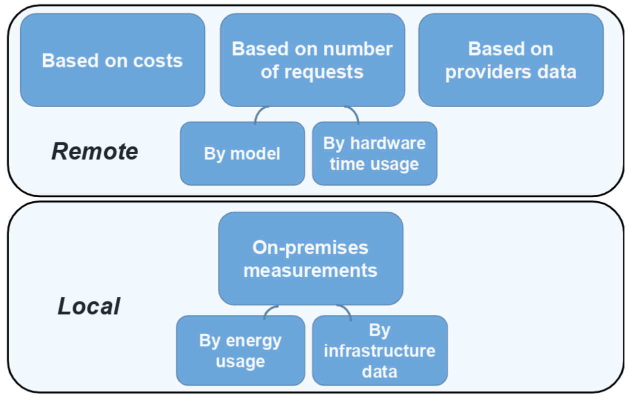

Figure 2. Project methodology

4.1. Based on Costs

The cost-based method is based on the maximum usage threshold for an AI service before its operational costs become unprofitable for its provider. Equation 1 is used, requiring the variables presented in Table 3. The flat rate depends on the provider and on the subscription tier.

$$I_{costs} = \frac{B \times M_e}{30 \times 24 \times C_e \times L} \tag{1}$$

Table 3. Variables and results of the cost-based method calculation.

Variable	Description	Units
B	Flat rate or pay-per-use costs	$costs/month$
L	Rate limit of the service	req/h
C_e	Energy costs	$costs/kWh$
M_e	Energy mix	gCO_2eq/kWh
I_{costs}	Environmental impact	gCO_2eq/req

4.2. Based on Requests Count

The method based on the number of requests requires a certain amount of data, as well as the total number of requests made over a given period within the assessed organisation. Depending on the data availability, this method can be calculated by relying either on the models or on the execution time.

4.2.1. Requests Count by Model

The environmental impact of requests through their quantity is calculated directly by the *EcoLogits*[1] tool, which provides a precise calculator for this scenario. Table 4 shows the variables to specify on its interface. Each of those variable

[1]EcoLogits address: `https://huggingface.co/spaces/genai-impact/ecologits-calculator`

have a strong effect on the consumed energy (Samsi et al., 2023).

Table 4. Variables and results of the request-based method calculation, as per models.

Variable	**Description**	**Unit**
Provider / Model	Used model	Enum
Parameters	Active & total, facultative	-
# tokens (input)	Input length	Enum
# tokens (output)	Expected output length	-
Data centre	PUE, WUE, location	-, -, Country
Energy mix	Emission for energy prod.	$kgCO_2eq/kWh$
Energy	Energy consumption	mWh
GHG	Effect on global warming	$mgCO_2eq$
Abiotic	use of metal & minerals	$\mu gSbeq$
Primary energy	Use of natural energy	kJ
Water	Water consumption	Ml

4.2.2. Requests Count by Time

The environmental impact of requests through time is calculated directly using the *Green Algorithms*[2] tool. Table 5 shows the variables to specify on its interface. Those variables can be configured for training phases, inferences phases, or both.

Table 5. Variables and results of the request-based method calculation, as per execution time.

Variable	**Description**	**Unit**
Reporting period	Overall period considered	time
Runtime	Task runtime, or continuous	time
Components	CPUs/GPUs: number & model	-
Usage factor	Real usage of CPUs/GPUs	%
Memory	Available memory	GB
Platform	Where is the task executed	Enum
Region	For remote data centre	Enum
Provider	For a remote data centre	Enum
PUE	For local data centre	Ratio
R&D training†	Include R&D overhead	-
Retraining*	Include all training iterations	-
Mult. factor*	To consider n inferences	-
Energy	Consumed energy	kWh
Carbon footprint	Environmental impact	$kgCO_2eq$

† = for training phase only, * = for inference phase only.

4.3. Based on providers data

The provider-based method relies on data provided by AI providers through their control panels, generated with the statistics of their services utilisation.

If providers have adopted the Greenhouse Protocol (GHG), the provided calculations can be be directly used as environmental impact. If providers only share information on the amount of requests counted by their systems, data must be combined with the request-based method presented in Subsection 4.2. If providers do not release any usage data, this method can not be used.

4.4. On-premises

Complementary to the previous three methods presented for remote services, the on-premises method directly measures local resources thanks to high-quality, specific knowledge and control of the infrastructure. Depending on the context, this method can be applied either when the overall energy consumed by the AI infrastructure is known, or when details about used hardware and infrastructure are known.

4.4.1. Known Energy Usage by AI

When the energy consumption of the assessed infrastructure is known, including overhead represented by the Power Usage Effectiveness (PUE), the straight-forward Equation 2 is used, with its variables presented in Table 6.

$$I_{usage} = \frac{E \times M_e}{1000} \tag{2}$$

Table 6. Variables and results used to calculate impact of on-premise infrastructure overall energy consumption.

Variable	**Description**	**Unit**
E	Total energy	kWh
M_e	Energy mix	$kgCO_2eq$
I_{usage}	Environmental impact	$kgCO_2eq$

4.4.2. Known Hardware and Infrastructure Data

When data about the infrastructure utilisation is known, the AI environmental impact is calculated using Equation 3 (Boavizta, 2026), with its required variables presented in Table 7. The Equation can be used for different type of equipment (rack, server, GPU, CPU, et cetera), and should therefore be repeated for each considered equipment.

$$I_{usage} = \frac{P_{hw} \times t_{usage} \times u}{1000} \times PUE_{dc} \times M_e \tag{3}$$

[2]Green Algorithms address: `https://calculator.green-algorithms.org/ai`

Table 7. Variables and results of the on-premises method, for AI models' usage phase.

Variable	Description	Unit
P_{hw}	Equipment average power	$Watts$
t_{usage}	Equipment usage time	h
u	Average equipment usage rate	$\%$
PUE_{dc}	Data centre PUE	-
M_e	Energy mix	gCO_2eq/kWh
I_{usage}	Usage environmental impact	$kgCO_2eq$

4.5. Other Calculations

The following Equations are useful for specific calculations and situations, and can be integrated into the overall impact estimation if desired.

4.5.1. AI Models Training Phase

The training phase of AI models has been determined as negligible, because it is most of the time drowned out by its usage phase (Strubell et al., 2019; Wu et al., 2022; Barr, 2022; International Energy Agency, 2025). However, some contexts, such as models trained for internal usage, might represent a higher relative impact compared to its usage, as observed in the interviews presented in Section 2. To this end, Equation 4 (Patterson et al., 2021) is used, requiring the variables presented in Table 8.

$$I_{train} = \frac{t_{train} \times N_p \times P_p \times PUE_{dc}}{1000} \times M_e \qquad (4)$$

Table 8. Variables and results of the on-premises method, for AI models' training phase.

Variable	Description	Unit
t_{train}	Total training time	h
N_p	Amount of chips	-
P_p	Average chip power	W
PUE_{dc}	Data centre PUE	-
M_e	Energy mix	$kgCO_2eq/kWh$
I_{train}	Train environmental impact	$kgCO_2eq$

4.5.2. Embedded Carbon

Embedded carbon represents the quantity of carbon emitted by the fabrication phase of an equipment. Despite being out of scope of the methodology, its inclusion for on-premises infrastructure might be relevant. Equation 5 (Boavizta, 2025; Morand et al., 2024) allows to obtain such environmental impact, with respect to a depreciation period, and requires the variables presented in Table 9. The Equation follows the attributional LCA approach, quantifying the environmental burdens attributed to a specific product. It can be used for a high variety of equipment (servers, GPUs, CPUs, et cetera), so long as data exists for them.

$$I_{embed} = I_{total} * \frac{t_{usage}}{t_{life}} \qquad (5)$$

Table 9. Variables and results used to calculate embedded impacts of equipment.

Variable	Description	Unit
I_{total}	Total emissions of manufacturing	$kgCO_2eq$
t_{usage}	Usage time	years
t_{life}	Estimated or actual lifespan	years
I_{embed}	Environmental impact	$kgCO_2eq$

4.5.3. Average Power Consumption

Equation 6 (International Energy Agency, 2025) can be used to approximate the average electricity consumption of an equipment, requiring the variables listed in Table 10.

$$P_{avg} = (P_{max} - P_{idle}) \times u + P_{idle} \qquad (6)$$

Table 10. Variables and results used to calculate the average power consumption of an equipment

Variable	Description	Unit
P_{max}	Equipment maximal power consumption	W
P_{idle}	Idled equipment power consumption	W
u	Equipment utilisation rate	$\%$
P_{avg}	Average electricity consumption	W

5. Case Study

We applied our methodology on The University of Lausanne (UNIL), a large-size organisation of more than 20,000 individuals, testing whether our work would allow to assess their AI environmental impact. This assessment has been conducted as a real case study.

Consistent with the challenges typical of large-scale organisations, data acquisition was hindered by restricted access to critical information: data inherent to AI-related spendings and usages was decentralised across departments, therefore complicating and delaying the data collection phase.

We also encountered a shortage of data necessary for calculating the provider-based method, resulting in its abandonment. The reason is mainly due to the decentralised nature of the organisation. Additionally, despite having a non-disclosure agreement with the organisation, a certain amount of data could not be collected for privacy and security reasons.

To implement our request-based method, we utilised data from a survey completed by a substantial number of organisational members. This survey provides a comprehensive assessment of the state of AI adoption within the UNIL at the time of the study. By synthesising data across several key metrics, the study establishes a baseline for understanding how AI services are internalised into workflows. The survey includes considerations about members' roles, their opinions on AI, the providers they use, and many other questions.

No statistical validation was provided with the study. However, 1,891 answers have been collected out of the approximative 21,000 UNIL members. Such figures provide a good basis, but do not ensure any strong representativeness.

Despite the survey large coverage, weak hypotheses had to be made to estimate the length of the requests, the usage frequency, and the used models. Those hypotheses have been established based on all possible trustable grounds and on all known data.

The survey covered all major types of common AI usages: chatbots, translation and image generation constitute almost all use cases, with the chatbot being by far the most common usage, followed by translation. Our methodology does not provide estimations for image generation.

The overall environmental impact we calculated for remote AI services of our study case are presented in Table 11, categorised by the type of AI services.

Table 11. Results of our case study environmental impact, for remote AI services. All three methods are shown, for each type of AI usage. N/D = No Data, N/A = Not Applicable.

	Impact (tCO2eq/year)		
Type	**Request count**	**Cost**	**Providers**
Chatbot	98.55	25.7	N/D
Translations	2.94	13.6	N/D
Image	N/A	N/A	N/D
Total	101.49	39.3	N/D

For the on-premise method, we had few hypotheses to make as we had access to most of the information we needed. The UNIL did most of its calculations through servers of two data centres it possesses or rents. We therefore obtained results that can be considered as reliable: we calculated an impact of 9.01 tCO_2eq for the 2024 year. This amount includes AI models training phases, as such activities are central in the studied organisation.

Table 12 shows the impacts calculated by each method, both remote than on-premises. We can observe that comparable impacts, which are the requests-based and cost-based ones, are separated by a factor of 2.58. While we could argue that this factor is far away from an order of magnitude, we could also state that those two amounts are not as close as wished. This table also shows attributional impacts to each member of the UNIL, reflecting average yearly emissions between 2.42 and 5.52 of $kgCO_2eq$ per person when considering remote and on-premises scopes.

Table 12. Comparison of calculated impacts for all the methods, alongside indication of impact per person (PP).

Method	**Overall impact** tCO2eq/year	**Impact pp.** kgCO2eq/year
# of requests	101.49	5.07
Costs	39.3	1.97
Suppliers	N/A	N/A
On-premises	9.01	0.45

6. Discussion

The project discussion mainly covers three parts.

6.1. Organisations Interest

The interviews show that the majority of surveyed organisations does not have a formal sustainability policy, but some are implementing similar measures. Respondents show little interest in an AI environmental impact audit solution, an interest highly correlated to its cost.

Overall, the interest of both companies and their customers in the environmental impact of the AI services they use is rather low. Moreover, many of the concepts, figures and facts communicated by the interviewers to the respondents were not understood by the latter, suggesting that opinions could change with a better awareness of the subject. However, there is interest in performing audits on the subject if the prices are reasonable.

6.2. Methodology Issues and Limitations

Calculations of impacts differ significantly depending on the method used, highlighting a lack of accuracy in the results obtained when using request-based or cost-based methods.

The methodology can be difficult to apply on large and complex organisations, especially when data is not centralised. Data confidentiality concerns can also arise, with assessors not having access to certain data.

Large organisations, such as the subject of this case study, typically conduct surveys similar to the one utilised here. In most instances, however, these instruments yield primarily qualitative insights. Regarding the request-based method, more precise tracking, for instance via network traffic analysis, could enhance accuracy, though it would introduce substantial technical and implementation complexities.

Regarding the different methods, their results can be hard to compare if there is missing data for one or another, although they can be complementary to each other in some situations. This highlights the fact that reliable hypotheses are hard to establish when needed data is missing.

As outlined in the assumptions presented in Point 3.2.2, we defined the variable L of Equation 1 as the rate limit of AI services to set an upper bound estimate of the costs where providers can no longer cover their operating expenses. Defining L differently, for instance as an average number of requests, could be considered. However, it would necessitate additional efforts to identify when services reach a saturated state.

Due to the lack of universal, reliable tools and calculations (Corley, 2026), our methodology is unfortunately dependent on external works. Problems can arise when updates or discontinuations are decided by maintainers and governance bodies. As an example, the request-based results presented in Section 5 have been presented to the assessed organisation in-between a refreshment of the EcoLogits methodology, resulting in different impact figures. The two versions have been presented to stakeholders, without anyone noticing any difference. While showing that such updates can totally change subsequent operational decisions, it also indicates that calculated impacts are not challenged by stakeholders due to a lack of knowledge.

For now, the methodology only supports AI services generating texts: multimodal models are currently unsupported, despite image and video generations gaining interest. Integrating such impacts would bring an extra value to the methodology. Another improvement consists of developing a user-intuitive manner to embrace and to use the methodology, including a comprehensive list of data that must be collected by assessed organisations.

6.3. Sectorial Evolution

One big issue that almost always arises when assessing ICT environmental impacts is to obtain access to high-quality, comprehensive data. To fill in this gap, AI services/models providers must share their internal measurements and information, despite revealing potential harmful strategic insights. Such paradigm shift could be imposed by political motivation: projects such as the AI Act, currently in discussion by the European Union, is a step forward such transparency. However, this Act should impose stricter policies on providers information disclosure to maximise the quantity of available data for the research community (Alder et al., 2025; Hacker, 2024; Ebert et al., 2025).

Overall, the sustainable ICT, Green computing, and related communities must converge into an effort to conceive precise, complete, exhaustive modelling of all components involved in providing AI services. Indeed, estimating impacts of a system while relying on incomplete and approximative models only results in increased imprecisions and noises accumulated through the components layering, transforming any conclusion into inaccurate figures. Moreover, the disparity of infrastructures, AI models, and hardware makes it difficult to compare such results. We hope that collective efforts will quickly be undertaken by the research community and providers to improve environmental impacts modelling.

Moreover, when one relies on existing and imprecise methods to either perform an environmental impact assessment or to define a new methodology, the successive accumulations of noises and imprecisions stack up due to the lack of acknowledged ground truth. Such conditions imply a strong dependence between presented results and the quality of used methods, resulting in a situation of *garbage-in, garbage-out*. In this case, researcher should integrate a disclaimer alongside their takeaways, avoiding any laundering being made by covering up uncertainties.

7. Conclusion

Despite an expanding body of research offering tools to measure the impacts of AI services, computations, or hardware, a standardised methodology for organisational assessment remains elusive. This study proposes an initial framework; however, while it facilitates preliminary estimates, several limitations prohibit the acquisition of definitive figures.

The primary obstacle to precise estimation is provider opacity regarding models, services, and infrastructure configurations and data. Should such data become available, refined modelling would enable high-fidelity environmental impact assessments. Furthermore, organisational complexity complicates the tracking of AI activities across usage and cost dimensions. Ultimately, this work underscores the need for the research community to advocate for provider transparency, prioritise low-level component modelling, and acknowledge the current lack of reliable methodologies.

Acknowledgements

This work has been supported by the HES-SO rectorate through research grant *Evaluation des Impacts Environnementaux de l'Intelligence Artificielle en Entreprise (EIEIAE)* - 134202/RI-PSTRAT24-01.

Limitations

Limitation of this work have already been mentioned, especially in the State of the Art, Section 3, and in the Discussion, Section 6. Moreover, one of the key takeaways of the *EIEIAE* project underlines the current limits of its own research question.

Ethical Statement

This work does not raise any ethical issues. The tools and data mentioned in our methodology are all publicly available and free to used if cited. If any, all licenses terms have been respected.

All information deducted and extracted by the interviews have been anonymised, and no personal data have been processed.

Researchers and collaborators of the *EIEIAE* project have travelled exclusively by train; besides computational power of computers for office tasks, no significant emissions of CO_2 have been emitted.

References

Admin, E. The Lifecycle of servers: Knowing when and how to decommission - ERI, 1 2024. URL https://eridirect.com/blog/2024/01/the-lifecycle-of-servers-knowing-when-and-how-to-decommission/.

Alder, N., Ebert, K., Herbrich, R., and Hacker, P. Ai, climate, and transparency: Operationalising and improving the ai act. *Journal of European Consumer and Market Law*, pp. 166–170, 2025. ISSN 2364-4710. URL https://kluwerlawonline.com/journalarticle/Journal+of+European+Consumer+and+Market+Law/14.2/EuCML2025026.

Amazon Web Services, Inc. Carbon footprint reporting – customer carbon footprint tool – amazon web services, 2 2026. URL https://aws.amazon.com/aws-cost-management/aws-customer-carbon-footprint-tool/.

AWS Inc. AWS Cloud - Amazon Sustainability, 2025. URL https://sustainability.aboutamazon.com/products-services/aws-cloud.

Barr, J. Amazon EC2 Update – Inf1 Instances with AWS Inferentia Chips for High Performance Cost-Effective Inferencing — Amazon Web Services, 11 2022. URL https://aws.amazon.com/blogs/aws/amazon-ec2-update-inf1-instances-with-aws-inferentia-chips-for-high-performance-cost-effective-inferencing/.

Berthelot, A., Caron, E., Jay, M., and Lefèvre, L. Estimating the environmental impact of generative-ai services using an lca-based methodology. *Procedia CIRP*, 122:707–712, 2024. ISSN 2212-8271. doi: https://doi.org/10.1016/j.procir.2024.01.098. URL https://www.sciencedirect.com/science/article/pii/S2212827124001173. 31st CIRP Conference on Life Cycle Engineering.

Boavizta. Get the impacts of a server - boavizta api documentation, 2025. URL https://doc.api.boavizta.org/getting_started/single_server/.

Boavizta. NocoDB, 2025. URL https://db.boavizta.org/dashboard/#/base/e3ba7aca-a9a7-4984-ad8d-8949ba47e305/md_kvka0tcnh8dyrw.

Boavizta. Embedded impacts - Boavizta API documentation, 2025. URL https://doc.api.boavizta.org/Explanations/embedded_methodology/.

Boavizta. Usage methodology - Boavizta API documentation, 2026. URL https://doc.api.boavizta.org/Explanations/usage/usage/.

Borisruf. Ai emissions scenario generator, 2025. URL https://borisruf.github.io/carbon-footprint-modeling-tool/ai-scenarios.html.

CodeCarbon. Codecarbon - tracking carbon emissions from machine learning, 2025. URL https://codecarbon.io/. Accessed: 2026-01-15.

Corley, V. e. a. All LCA models are wrong. Are some of them useful? Towards open computational LCA in ICT. *In preparation*, 2026.

Donnellan, D., Lawrence, A., Bizo, D., Judge, P., O'Brien, J., Davis, J., Smolaks, M., Williams-George, J., and Weinschenk, R. Uptime institute global data center survey 2024. Technical report, Uptime Institute, 2024. URL https://datacenter.uptimeinstitute.com/rs/711-RIA-145/images/2024.GlobalDataCenterSurvey.Report.pdf.

Ebert, K., Alder, N., Herbrich, R., and Hacker, P. Ai, climate, and regulation: From data centers to the ai act, 2025. URL https://arxiv.org/abs/2410.06681.

EcoLogits. LLM inference - EcoLogits, 11 2025. URL https://ecologits.ai/latest/methodology/llm_inference/.

Ecologits. _ecologits - ecologits, 11 2025. URL https://ecologits.ai/latest/reference/_ecologits/.

Frymire, L. and Owen, D. The stock of computing power from nvidia chips is doubling every 10 months, 2025. URL https://epoch.ai/data-insights/nvidia-chip-production. Accessed: 2025-04-30.

Frymire, L., Snodin, B., and Owen, D. Leading ai chip designs are used for around 4 years in frontier training, 2025. URL https://epoch.ai/data-insights/gpu-frontier-lifespan. Accessed: 2025-04-30.

GeSI. Carbon trust (2017): Ict sector guidance, built on the ghg protocol product life cycle accounting and reporting standard, july 2017, 2017. URL https://ghgprotocol.org/sites/default/files/ghgp/GHGP-ICTSG%20-%20ALL%20Chapters.pdf.

Google Cloud. Power usage effectiveness – Google Data Centers, 2026a. URL https://datacenters.google/efficiency/.

Google Cloud. Carbon footprint, 2026b. URL https://cloud.google.com/carbon-footprint.

Hacker, P. Sustainable ai regulation. *Common Market Law Review*, pp. 345–386, 2024. ISSN 0165-0750. URL https://kluwerlawonline.com/journalarticle/Common+Market+Law+Review/61.2/COLA2024025.

Hugging Face. Ai energy score leaderboard - a hugging face space by aienergyscore, 2025. URL https://huggingface.co/spaces/AIEnergyScore/Leaderboard.

Hummel, K. and Jobst, D. An overview of corporate sustainability reporting legislation in the european union. *Accounting in Europe*, 21(3):320–355, 2024. doi: 10.1080/17449480.2024.2312145. URL https://doi.org/10.1080/17449480.2024.2312145.

International Energy Agency. Energy and AI – Analysis - IEA, 4 2025. URL https://www.iea.org/reports/energy-and-ai.

Iwata, H. and Okada, K. Greenhouse gas emissions and the role of the Kyoto Protocol. *Environmental Economics and Policy Studies*, 16(4):325–342, 9 2012. doi: 10.1007/s10018-012-0047-1. URL https://doi.org/10.1007/s10018-012-0047-1.

Lacoste, A., Luccioni, A., Schmidt, V., and Dandres, T. Quantifying the carbon emissions of machine learning, 2019. URL https://arxiv.org/abs/1910.09700.

Lannelongue, L., Grealey, J., and Inouye, M. Green algorithms: quantifying the carbon footprint of computation. *Advanced science*, 8(12):2100707, 2021.

Lannelongue, L., Grealey, J., and Inouye, M. Green algorithms - ai view, 2026. URL https://calculator.green-algorithms.org/ai.

McKinsey. Investing in the rising data center economy, January 2023. URL https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/investing-in-the-rising-data-center-economy.

Microsoft. Emissions impact dashboard — microsoft sustainability, 2026. URL https://www.microsoft.com/en-us/sustainability/emissions-impact-dashboard.

ML.ENERGY. Ml.energy leaderboard, 12 2025. URL https://ml.energy/leaderboard/.

Morand, C., Ligozat, A.-L., and Névéol, A. Mlca: A tool for machine learning life cycle assessment. In *2024 10th International Conference on ICT for Sustainability (ICT4S)*, pp. 227–238, 2024. doi: 10.1109/ICT4S64576.2024.00031.

Mytton, D. Assessing the suitability of the Greenhouse Gas Protocol for calculation of emissions from public cloud computing workloads. *Journal of Cloud Computing Advances Systems and Applications*, 9(1), 8 2020. doi: 10.1186/s13677-020-00185-8. URL https://doi.org/10.1186/s13677-020-00185-8.

Nilvér, K. The carbon footprint of servers – goclimate blog, 2019. URL https://www.goclimate.com/blog/the-carbon-footprint-of-servers/.

Nordic Computer. What is the effect of datacenter hardware on the climate?, 2022. URL https://nordiccomputer.com/resources/what-is-the-effect-of-datacenter-hardware-on-the-climate.

Nover, S. Will AI companies ever be profitable?, 11 2024. URL https://www.gzeromedia.com/gzero-ai/will-ai-companies-ever-be-profitable.

Optimum. Llm-perf leaderboard - a hugging face space by optimum, 2023. URL https://huggingface.co/spaces/optimum/llm-perf-leaderboard.

OVH. Ovhcloud's carbon calculator methodology, 11 2023. URL https://corporate.ovhcloud.com/sites/default/files/2023-11/methodo_carboncalc-en.pdf.

Patterson, D., Gonzalez, J., Le, Q., Liang, C., Munguia, L.-M., Rothchild, D., So, D., Texier, M., and Dean, J. Carbon emissions and large neural network training, 2021. URL https://arxiv.org/abs/2104.10350.

Pineda, J., Konikoff, J., Rajendran, S., Ahlawat, P., and Izaret, J.-M. GenAI needs pricing strategies to match its potential, 5 2024. URL https://www.bcg.com/publications/2024/genai-needs-pricing-strategies-to-match-its-potential.

Retegui Schiettekatte, A. F. Environmental impact of producing and operating server infrastructure : A systematic literature review, 2021.

Rincé, S., Banse, A., Defour, V., and Jean-Pierre, C. Ecologits calculator, 2025. URL https://huggingface.co/spaces/genai-impact/ecologits-calculator.

Russell, S. TechInsights platform, 2 2025. URL https://library.techinsights.com/public/hg-asset/e95a4878-6d15-4bf5-a970-02323d482a38.

Safaei, A. Greenhouse gas protocol. *Sector Toolsets for Iron and Steel-Guidance Document*, pp. 1–12, 2011.

Samsi, S., Zhao, D., McDonald, J., Li, B., Michaleas, A., Jones, M., Bergeron, W., Kepner, J., Tiwari, D., and Gadepally, V. From words to watts: Benchmarking the energy costs of large language model inference. In *2023 IEEE High Performance Extreme Computing Conference (HPEC)*, pp. 1–9, 2023. doi: 10.1109/HPEC58863.2023.10363447.

Shilov, A. Datacenter GPU service life can be surprisingly short - only one to three years is expected according to unnamed Google architect, 10 2024. URL https://www.tomshardware.com/pc-components/gpus/datacenter-gpu-service-life-can-be-surprisingly-short-only-one-to-three-years-is-expected-according-to-unnamed-google-architect.

Strubell, E., Ganesh, A., and McCallum, A. Energy and policy considerations for deep learning in NLP. In Korhonen, A., Traum, D., and Màrquez, L. (eds.), *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 3645–3650, Florence, Italy, July 2019. Association for Computational Linguistics. doi: 10.18653/v1/P19-1355. URL https://aclanthology.org/P19-1355/.

The Shift Project. Intelligence artificielle, données, calculs : Quelle infrastructures dans un monde décarboné ?, 2025. URL https://theshiftproject.org/publications/intelligence-artificielle-centres-de-donnees-rapport-final/.

Tomlinson, B., Black, R. W., Patterson, D. J., and Torrance, A. W. The carbon emissions of writing and illustrating are lower for AI than for humans. *Scientific Reports*, 14(1), 2 2024. doi: 10.1038/s41598-024-54271-x. URL https://doi.org/10.1038/s41598-024-54271-x.

Whitwam, R. Data center ai gpus may have extremely short lifespans, 10 2024. URL https://www.extremetech.com/computing/data-center-ai-gpus-may-have-extremely-short-lifespans. Accessed: 2025-04-30.

Wu, C.-J., Raghavendra, R., Gupta, U., Acun, B., Ardalani, N., Maeng, K., Chang, G., Aga, F., Huang, J., Bai, C., Gschwind, M., Gupta, A., Ott, M., Melnikov, A., Candido, S., Brooks, D., Chauhan, G., Lee, B., Lee, H.-H., Akyildiz, B., Balandat, M., Spisak, J., Jain, R., Rabbat, M., and Hazelwood, K. Sustainable ai: Environmental implications, challenges and opportunities. In Marculescu, D., Chi, Y., and Wu, C. (eds.), *Proceedings of Machine Learning and Systems*, volume 4, pp. 795–813, 2022. URL https://proceedings.mlsys.org/paper_files/paper/2022/file/462211f67c7d858f663355eff93b745e-Paper.pdf.

Zitron, E. There is no AI revolution, 2 2025. URL https://www.wheresyoured.at/wheres-the-money/.

A. Interview Questions

The interviews that have been conducted to identify the needs and opinions of organisations about AI environmental impacts were the following:

- Does your company have a corporate policy on sustainability or the environment?
- Are your company's sustainable actions implemented for internal or external reasons?
- Is sustainability a criterion for tenders?
- Have you ever heard of the CSRD?
- Do you have any idea what your carbon footprint is? Or has anyone ever mentioned this subject?
- Is environmental impact discussed with your customers?
- How many people are there in your company? And how many employees work as data scientists?
- Do IT costs account for 10% of your company's turnover?
- Do these IT costs include subscriptions to external services?
- What proportion of costs does AI represent?
- Would an environmental audit service for IT focused on AI, at CHF 5k, be of interest?

B. Details on EcoLogits

The EcoLogits calculator uses a database to make estimations for specific tasks provided IaaS models such as chatbots, for which we may not know the precise hardware used to perform the computations. It takes into account a broad number of parameters, including many details about the model, as well as the data centre configuration, for which we can enter the location to indicate its electric mix, its Power Usage Effectiveness (PUE) as well as its Water Usage Effectiveness (WUE). The calculator provides energy consumption, GHG emissions, as well as water and abiotic resources consumption for both embodied and usage contributions. Unlike the Python library, it's typically used for making estimations of the impacts of inference of the model before the start of a project in order to decide which models to use. Gen AI Impact regularly updates its models database, but often does not provide calculations for the most recent models.

C. Useful Data

The *EIEIAE* methodology equations may require data that is difficult to obtain. A certain amount of data have been found during our researches.

C.1. Embodied Carbon

The following content focuses on the attributional LCA approach.

In terms of embedded carbon, the fabrication of servers emit an average of 1525 $kgCO_2eq$ (Nordic Computer, 2022; Nilvér, 2019; Boavizta, 2025).

The embodied carbon of GPUs highly varies depending on their fabrication year (Russell, 2025) and models, with figures varying between 0.125 and 2 tCO_2eq (Tomlinson et al., 2024; Wu et al., 2022; Russell, 2025).

C.2. Usage Factor

Research suggests datacenter utilisation ratios of 50%, while others suggest 70% (Retegui Schiettekatte, 2021). The IEA puts forward figures of 20% pure utilisation for small data centres, which are less efficiently loaded, and 50% utilisation for hyperscale data centres (International Energy Agency, 2025).

GPUs usage factor depend mainly on the nature of the tasks: training phases can take up to 95%, and often around 80% (Berthelot et al., 2024). The inference portion accounts for around 40% (Berthelot et al., 2024).

C.3. Lifespans

The lifespan of servers depends greatly on how they are used and maintained. Online, people talk about periods ranging from 2 to 20 years, with an average of approximately 10 years (Admin, 2024). A meta-analysis determined that servers remain operational for between 3 and 4 years, according to 4 papers, and are then replaced (Retegui Schiettekatte, 2021). The IEA puts forward a figure of 5 years (International Energy Agency, 2025). The EcoLogits methodology has set the equipment usage period at 5 years (EcoLogits, 2025). We can conclude that an average period of 4.5 years can be considered.

The lifespan of GPUs is influenced the increasingly needs of computing power for AI models (Frymire & Owen, 2025), making GPU replacement more likely when acceleration of research and development phases is wished. For chips used to train the most advanced and heavy models, the average useful life is 3.9 years (Frymire et al., 2025). According to a Google engineer, the company uses their GPUs for an average of 3 years (Whitwam, 2024). Another interview, also with a Google employee, suggests that usage can last

3 years, but can drop to as little as one year (Shilov, 2024). However, it is important to note that these sources are not verifiable, as they are anonymous.

C.4. Energy Mix

One of the most widely recognised sources of energy mix data is the *Electricity Maps*[3] project, which is referenced by many researchers.

C.5. PUE

The Power usage effectiveness of datacenters is not always available, especially for remote services. However, estimations have been, with a worldwide ratio of 1.56 in 2024 (Donnellan et al., 2024). However, this ratio highly depends on the provider, and its wealth facilatating investments: for example, Amazon Web Services (AWS) reported an average PUE of 1.15 in 2024, with their best performing region being Europe (1.04) (AWS Inc., 2025). Google Cloud also reported low PUE figures, with a ratio of 1.08 for 2024 (Google Cloud, 2026a).

C.6. Occupation of AI Tasks in Datacenters

Indications show that most of the energy demand dedicated to AI comes from datacenters, under the form of both cloud and hyperscaler (International Energy Agency, 2025). Given the fact that it is very difficult to distinguish between resources used for AI and those used for other tasks in data centres, the IEA proposes a proxy, namely the electricity consumption of GPU-accelerated servers, i.e. those specialised for AI-related tasks. These specialised servers account for 24% of electricity consumption within a data centre, all servers combined, and 15% of a data centre's total consumption in 2024 (International Energy Agency, 2025).

[3] Electricity Maps address: `https://app.electricitymaps.com`