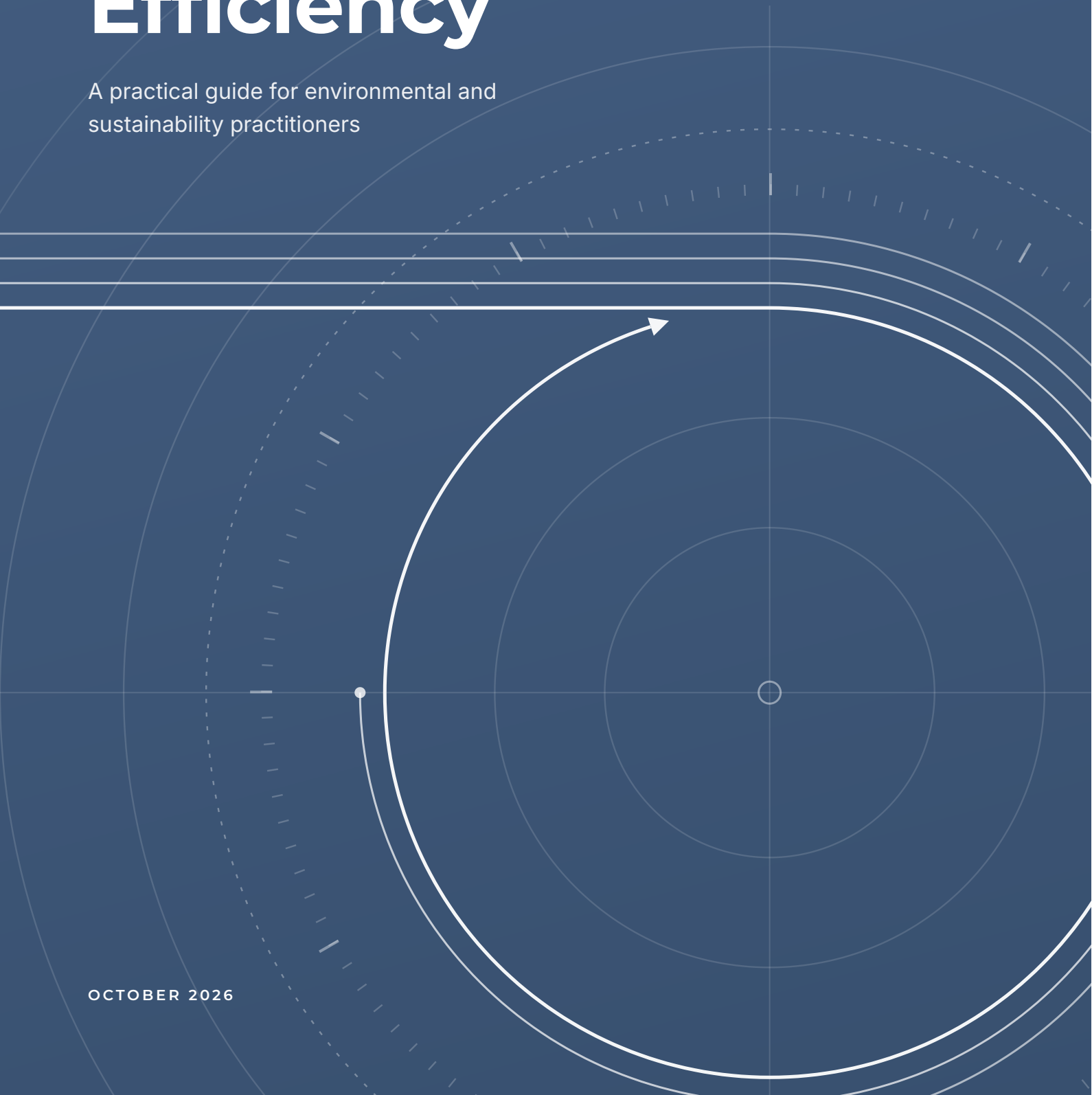
COMPASS WORKS

PRACTITIONER GUIDE

Managing AI Resource Efficiency

A practical guide for environmental and sustainability practitioners



OCTOBER 2026

Contents

01	AI is entering practice	03
02	Why AI resource efficiency warrants management attention	04
	The management approach: the AI resource efficiency cycle	**05**
03	Establish governance and accountability	06
04	Manage the demand	09
05	Check and measure	12
06	Turning evidence into decisions	15
07	Closing	17
	Methodology	18
	References	19

AI is entering practice

As discussion of AI's environmental impact grows among practitioners, the practical question is shifting from awareness to management.

In the space of a year or so, AI has been built into the email, documents and spreadsheets organisations already use. By mid-2026, AI assistants were included in the standard business packages of both Google Workspace and Microsoft 365.[1] ESG teams are using it too: to draft disclosures, summarise supplier policies and pull figures out of utility invoices. Similar capabilities are also appearing inside the specialist platforms and enterprise systems through which organisations manage everyday work.

That convenience rests on physical infrastructure. AI requests are processed in data centres that draw electricity, rely on cooling that can consume water, and run on specialised hardware that has to be made, housed and eventually replaced. How much this matters depends on where and how the work is done. In our view, that makes the environmental question wider than carbon: energy, water and hardware matter too.[2]

Data centres in numbers

WORLDWIDE · 2025
IEA ESTIMATE

485 TWh

of electricity used by data centres, up 17% in a year against 3% growth in electricity demand overall. Electricity use at AI-focused data centres grew about 50%.[3]

GREAT BRITAIN · 2024
OFFICIAL ESTIMATE

4.5 TWh

about 2% of electricity drawn from the grid, up 41% since 2020.[4]

GREAT BRITAIN · 2030
NESO FORECAST

Just over 20 TWh

a year. NESO describes the outlook as highly uncertain.[5]

These figures cover data centres as a whole, not AI alone. The 2024 estimate and 2030 forecast are produced on different bases and should not be directly compared.

Purpose of this guide

Professional discussion of AI covers both what it makes possible and the resources it relies on. For environmental practitioners, that second side is no longer someone else's subject. Rising use, and rising demand for the infrastructure behind it, give practitioners good reason to understand AI's resource use alongside the energy, water and materials they already work with. The point is not to create a separate environmental discipline around AI, but to bring its resource use into the same practical conversation about demand, efficiency and evidence.

Knowing that AI is valuable, or that it has an environmental cost, is a start. It does not, by itself, tell an organisation what to do about the demand it creates. Environmental practitioners will recognise the pattern: knowing a building's energy use is not the same as making it efficient.

It is therefore time to look at AI's resource use more seriously, and to move from awareness to something practitioners can start to act on. That is why this guide was put together.

NOTES

1 These sources show that AI is included in standard plans, not how much it is used. Google (2026a); Microsoft (2026).
2 EU data-centre reporting, for example, covers energy, water and waste heat separately. Regulation (EU) 2024/1364.
3 An IEA estimate. The IEA does not give AI's share of the total. IEA (2026).
4 Official statistics in development. The figure excludes data centres that organisations run in-house. DESNZ (2026).
5 NESO builds its estimates from connected capacity, not metered consumption, so they are not comparable with the official estimate above. NESO (2026).

Harder to manage than it looks

AI resource efficiency asks three plain questions.

1. What outcome is actually needed?
2. What computational resources are being used to deliver it?
3. Is that use proportionate to the outcome?

Why it is harder to manage

For AI, those questions are harder to answer than for many of the resources an environmental team already manages. Demand can start almost anywhere. Because AI now comes built into the software people already use, a request can be made by any employee, inside a tool the organisation has already paid for, or by an automated workflow that runs without anyone pressing a button. That makes AI resource use an organisational question, not just a procurement one.[1]

Nor is the demand uniform. Two requests that look alike to the person making them can differ in computational intensity by orders of magnitude, depending on the task, the model and how the service is set up.[2]

Most of it is also hard to see. What organisations see are licences, prompts, credits and invoices. Environmental data from providers, where it exists, generally stops at the provider's own infrastructure and rarely reaches the level of an individual organisation's tasks.[3] The methods and disclosures that might change this are still moving.[4]

And the technical work is a specialism. An environmental practitioner can set expectations, ask for evidence and judge whether management is credible, but tuning models or infrastructure is engineering work. That calls for a division of labour, not a new competence for ESG.[5]

Its organisational and technical characteristics make AI resource efficiency harder to manage than most familiar environmental aspects. The underlying management challenge, though, is a familiar one.

How much attention is proportionate?

None of this casts doubt on whether the impacts are real. The uncertainty is about what one organisation can see and attribute to itself. Where that picture is incomplete, the answer is proportionate management, not waiting until the impacts can be fully quantified.

AI's environmental relevance at system level does not make its use significant for every organisation. Significance is judged organisation by organisation. In our view, it should set how much management an organisation applies, not whether the basic principles of resource efficiency apply at all. The practical question is therefore not whether to act, but what level of attention is proportionate.[6]

A judgement, not a calculation[7]

Scale and pattern of AI uses across the organisation	×	Reason for environmental concern	=	Level of management response beyond a yes or no decision

NOTES

1 Most EU enterprises that use AI buy it ready-made. Eurostat (2026).

2 Published measurements show large variation across tasks and models, but do not represent the footprint of a particular hosted service. Luccioni, Jernite and Strubell (2024); Dauner and Socher (2025); Chung et al. (2025).

3 Providers' customer carbon tools allocate emissions within the provider's own cloud, not through another company's software. AWS (2025).

4 ISO/IEC 30134-2 (2026) and ITU-T L.1450 (2025), for example, were both revised in this period.

5 Skills frameworks describe typical professional scope, not a limit on individuals. SFIA Foundation (2024); ISEP (n.d.).

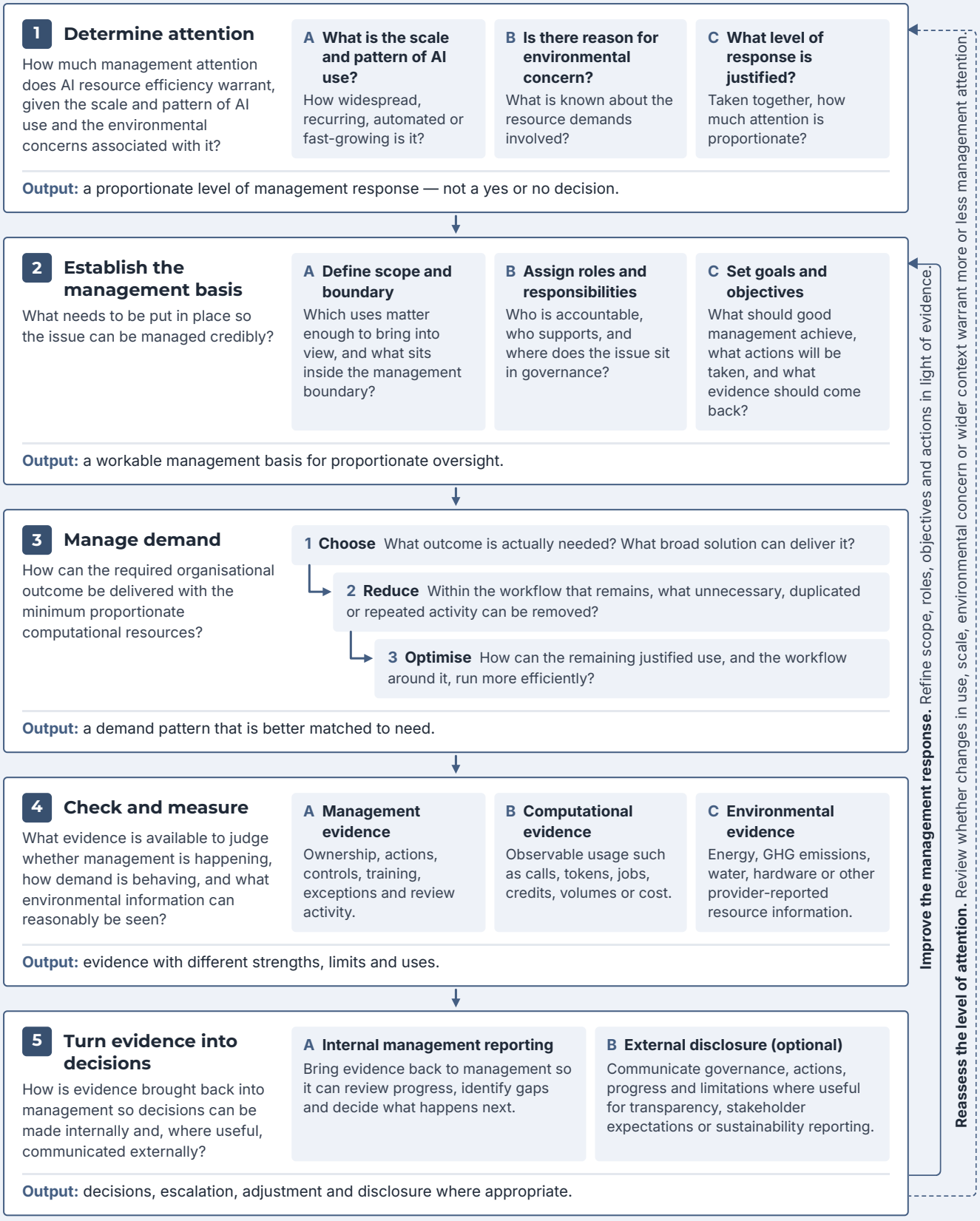
6 None of the standards or regulations reviewed for this guide treats AI use as material for every organisation. EMAS, Annex I; ESRS 1 (2023).

7 The first line follows EMAS, Annex I. The rest is our own extension.

Managing AI resource efficiency

A practical management cycle

The rest of this guide follows a practical management cycle. It starts by deciding how much attention AI resource efficiency warrants, then works through the management basis, demand decisions, evidence and reporting. The cycle is iterative: evidence should improve the management response and, where circumstances change, prompt a fresh view of how much attention is needed. The figure below provides the roadmap for the rest of the guide.



What needs to be managed?

Four things need to be in place

Deciding how much attention AI resource efficiency deserves only matters if something is then in place to act on it. That does not require a new governance system. These four things will be familiar from any management system.[1] Together, they provide the minimum management basis from which demand can be challenged, evidence can be gathered and decisions can be reviewed. How much formality they need follows from the level of attention already decided. A modest, stable use may need little more than a named owner, a clear intent and a line in an existing forum; significant or fast-growing use warrants more.

1. **The uses that matter are in view**
   Which AI uses matter enough to manage?
2. **Someone is accountable**
   Who is accountable, and who does the work?
3. **Set goals and objectives**
   What should good management achieve, and what will be done?
4. **Evidence comes back**
   What returns to governance, what is it checked against, and when?

1. Bring the uses that matter into view

Define a scope and boundary. Nobody needs to inspect every prompt. Prioritising by significance is the ordinary form of risk-based management, and both sustainability reporting and GHG accounting accept proxies and estimates in place of exhaustive measurement.[2] The uses worth bringing into view are those that are significant, recurring, scaled, automated, strategically important or growing fast. When looking at one, it helps to remember that demand and inefficiency are not only properties of the technology. They can start with the activity itself, the workflow around it, or the tool chosen to deliver it.

Some of those uses will be easy to see and some will not. Work-related AI use extends well beyond centrally procured tools. Employee-reported use substantially exceeds business-reported adoption, much of it runs through freely available public tools, and in one international study (2025), only around two in five employees reported that their organisation had a generative AI policy.[3] Even approved tools may reveal little about the computation behind them.

PRACTITIONER NOTE

Governance can reach further than measurement.

The boundary of an organisation's AI demand may be wider than the boundary it can currently observe or measure.[4]

For instance, employees can use their own AI subscriptions to assist a work-related task. That changes the form of governance, not whether it happens. Where use cannot reasonably be measured, our view is that the proportionate response is behavioural. That means:

- clear expectations for appropriate use;
- a preference for approved tools where they exist; and
- bringing significant or recurring uses into a managed arrangement.

Organisation-wide guidance and training on appropriate AI use are one of the ways those expectations reach people.

NOTES

1 These four elements mirror ISO management system standards: scope, responsibilities, objectives and planning, and performance review. ISO (2023).
2 Reporting standards accept estimates where data is lacking; UK regulators should focus on priority risks. ESRS 1 (2023); Regulators' Code (2014).
3 From an international employee survey. The fieldwork predates this guide by over a year, so the figure has likely moved. Gillespie et al. (2025).
4 The EU AI Act works similarly: a model's known or estimated energy use must be documented, so estimates are accepted. Regulation (EU) 2024/1689.

Who owns it, and where does it sit?

2. Make sure someone is accountable

Accountability for an environmental outcome, and execution of the technical work that delivers it, can sit with different functions. Established management practice separates the two, and professional skills frameworks place technical optimisation outside the usual scope of environmental practice. In our view, environmental oversight therefore does not require ESG to own technical optimisation, procurement or operational execution.[5] Environmental functions already work this way elsewhere: they do not need to replace lighting systems themselves to establish expectations, accountability and assurance that energy efficiency is being managed.

ESG does not need to own the technology to ensure that its environmental implications have an owner.

Roles and responsibilities to consider

The roles below illustrate those an organisation may need to consider. How they are allocated is specific to each organisation, and in a smaller organisation one person may hold several of them.

ROLE	LIKELY CONTRIBUTION
Business owner	Defines the required outcome, its value, scale and acceptable performance, including whether the need exists at all
AI or technical owner	Selects, designs, configures and optimises the solution and workflow
IT, digital and procurement	Manages architecture, approved technologies, suppliers, contracts and available usage data, and carries environmental expectations into supplier relationships
ESG / environmental	Sets environmental expectations, identifies gaps, asks for evidence and escalates where necessary
Governance, risk or executive management	Gives the mandate, resolves ownership gaps and integrates the issue into wider oversight

Identify the right governance structure

With the relevant owners identified, the next question is how they are brought together. AI resource efficiency is not necessarily a standalone subject to be managed. It can often be integrated into governance the organisation already has, for example:

- existing AI-related governance, such as governance of AI use, responsible AI or AI ethics;
- IT or digital governance;
- environmental governance; or
- enterprise risk management.

While not a prerequisite, AI management system standards such as ISO/IEC 42001 use a common structure intended "to promote integration across disciplines". Which arrangement is lightest and still credible is the organisation's judgement.[6]

PRACTITIONER NOTE

When nobody owns it

The absence of governance is itself a governance finding. Internal-control practice expects authority and responsibility to be assigned, so an issue with no credible owner is a condition to report, not an obstacle to reporting.[7] In our view, ESG's role is then to:

- establish the evidence;
- raise the issue through an appropriate senior route; and
- seek a proportionate mandate.

It should neither absorb the issue as its default owner nor wait for formal governance to appear.

NOTES

5 Internal-control practice separates being accountable for an outcome from carrying out the work that delivers it. COSO (2013).

6 From ISO's explanation of why its management system standards share one structure. ISO/IEC 42001 is the standard for AI. ISO (n.d.).

7 Internal-control frameworks expect responsibility to be assigned and people to be held accountable. COSO (2013); Institute of Internal Auditors (2024).

Set direction and agree what comes back

3. Set goals and objectives

Naming an owner is not the same as managing the issue. The owner first needs to understand what is at stake in their use of AI, decide what good management would look like, and translate that into a small number of proportionate objectives and actions.[8] Those objectives need not be elaborate, and they need not be environmental targets. For most AI uses, the environmental footprint of the work cannot yet be traced reliably, so objectives will often concern behaviour, decisions or observable activity instead. What matters is that they are clear enough for someone to tell later whether they were met.

For an individual AI use, the actions will usually be demand decisions: whether the need calls for AI at all, what activity can be removed, and how justified use can run more efficiently. General AI governance can host resource-efficiency management, but it does not by itself show that resource demand is being managed proportionately. The distinction is between the governance structure that hosts the issue and the management response it produces.

EXAMPLE

Start a policy from its purpose

An organisation decides it needs group-level guidance on AI use. Before drafting anything, it has to decide what the guidance is for.

- Is it mainly about confidentiality and accuracy?
- Should it also steer people towards approved tools?
- Should it discourage many people producing the same output separately?
- Should it reserve the heaviest AI capabilities for work that needs them?

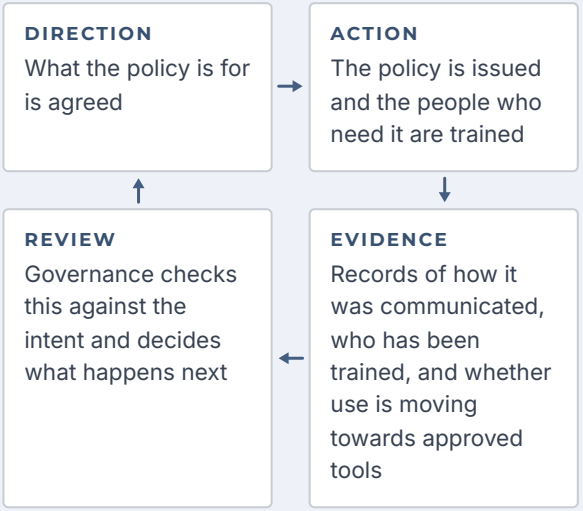
Only once that intent is agreed can the organisation settle the content, who needs training, and what would show the guidance is being followed. A policy that stopped at the first two questions would cover confidentiality, accuracy and approved tools without touching duplicated work or how much capability a task needs. That is no failing of such a policy. It answers a different question.[9]

4. Agree what comes back, and what it is checked against

Evidence is only meaningful against something. Accountable owners should return enough evidence for governance to judge whether the agreed objectives are being met and the agreed actions carried out. Governance should then review it and decide what happens next.

What would not help is a request for “AI emissions data” in the abstract. Agree at the outset what will come back, where it goes and roughly when. The evidence need not be elaborate. What matters is that the route exists before it is needed, and that everyone knows what the evidence is being compared with.

The policy example, followed through



NOTES

8 The standards ask for objectives to be measurable “if practicable”. ISO (2023).

9 AI governance frameworks do address resource use, chiefly as something to assess and document. NIST (2023).

The demand-management hierarchy

Choose, reduce, optimise

The central operational question is how the demand itself is managed. Doing that well means delivering the required organisational outcome using the minimum proportionate computational resources. That is not the same as using less AI. It means taking three separate decisions, each needing its own justification: **choose** the right solution, **reduce** what does not need to happen, and **optimise** what remains.[1]

1 CHOOSE an appropriate solution

What outcome is actually needed, and what kind of solution can deliver it? That includes whether AI is needed at all.

2 REDUCE unnecessary activity

In the workflow that remains, what unnecessary, duplicated or repeated processing can be removed, whatever the technology involved?

3 OPTIMISE justified use

How can the remaining justified AI use, and the workflow around it, run more efficiently?

EXAMPLE

Autumn Budget: one need, three decisions

Someone simply wants to know what the Budget says. An existing authoritative summary already meets the need.

No AI generation required.[2]

The Budget matters to the business. One team produces a structured, AI-assisted analysis of the provisions that affect the organisation and shares it.

Employees no longer repeat much the same work.

For that shared analysis, use the capability, reasoning depth, context and output the task needs.

Not the most computationally intensive option by default.[3]

The need and the broad solution class = Choose. Removing work that does not need to happen = Reduce. Doing the necessary work more lightly = Optimise.

Why the order matters

The order matters because each step changes what the next one works on. Technical optimisation matters, but it works on demand that has already been accepted, in a form that has already been chosen. Asking Choose and Reduce first helps ensure that optimisation effort is applied to demand that deserves it.

The Budget example shows why the steps are kept apart. Most of the demand in it disappears before any model is tuned. At Choose, part of the need turns out to require no generation at all. At Reduce, one shared analysis replaces many near-identical ones. Optimise still matters, but it is working on what is left. In practice the order is not always so clean, and at scale optimising may reveal that the workflow itself should be redesigned. A redesigned workflow is simply taken through the same three questions again.

Environmental practitioners will recognise the logic: prevention still comes before efficiency, as it does in managing waste or energy. We call the first step Choose, not Avoid, because the right decision may be to use AI. The aim is the appropriate solution, not the avoidance of AI.

NOTES

1 The IPCC distinguishes avoiding demand, shifting to more efficient provision and improving efficiency; EU waste law puts prevention first. The naming and the two-level split of prevention are our own. IPCC (2022); Directive 2008/98/EC.

2 "Sometimes, AI is not the best solution for your problem: it may be more easily solved with more established technologies." GDS (2025).

3 The size of the effect depends strongly on the task. Chung et al. (2025); Oviedo et al. (2025).

Using the hierarchy in practice

Two ways to use it

For ESG practitioners, the hierarchy has two distinct uses.

AS A COORDINATOR

Use the three questions to help other functions examine how AI demand is created and managed in the workflows they own.

ESG asks the questions and looks for the evidence. The function that owns the workflow makes the decisions.

AS A USER

Apply the same questions directly to the workflows ESG owns, shapes or influences.

Where ESG owns the problem a workflow exists to solve, it is well placed to judge what kind of AI use, if any, that problem warrants.

What can the organisation actually influence?

The levers available depend on where the organisation sits in the AI service chain. Where it controls the model call or workflow, it can act directly. Where AI is embedded in enterprise software, most technical levers sit with the supplier and are reached through expectations, procurement and supplier engagement. Provider or deployment choice matters only where it is controllable and supported by decision-useful resource information. The hosted services reviewed for this guide did not generally offer that.[4] Either way, ESG sets the environmental expectation and asks for evidence. It does not become the technical owner.

Why reduce activity if the physical saving is hard to see?

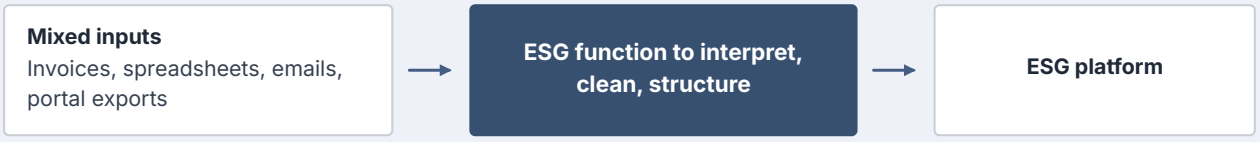
One avoided interaction does not necessarily produce an immediate, one-for-one physical saving, because hardware draws substantial power whether or not it is doing useful work.[5] That does not weaken the case for Reduce. Recurring unnecessary computation still adds demand and occupies capacity that is expensive in its own right, and it is easier to design out before it becomes embedded. The aim is to stop unnecessary activity becoming routine, not to claim a precise physical saving from each interaction avoided.

EXAMPLE PRACTITIONER SCENARIO

A new platform, the same messy inputs

Illustrative composite, not a client case. Where the lines fall will differ between organisations.

The situation. A company has implemented a new ESG platform, but the raw data feeding it is still messy. Invoices, spreadsheets, emails and other source information arrive in inconsistent formats and of varying quality.



Where the effort sits. The platform may improve the destination, but significant effort is still needed to interpret, clean and structure the inputs. In some cases that effort increases, because the new platform expects more structured or more complete data than before.

AI may therefore look like a plausible tool for dealing with the messy inputs. Before adding it, the same three decisions apply.

NOTES

4 For the hosted services reviewed (September 2026), customers had little or no control over where processing runs, and none published region-level environmental data for AI workloads.

5 Idle hardware still draws power, so average energy per interaction overstates what avoiding one saves. Most published per-query estimates assume proportionality, so this is a minority position. Lawrence Berkeley National Laboratory (2026).

The scenario, worked through

EXAMPLE PRACTITIONER SCENARIO

A new platform, the same messy inputs

1 CHOOSE *What changes*

Match each kind of input to the solution it actually needs.

Data already held in a structured system, such as a utility portal or finance system	→ direct integration or structured transfer
Invoices and forms in consistent, known layouts	→ conventional extraction or OCR
Known checks, such as units, date ranges, missing fields or implausible values	→ deterministic rules
Genuinely variable or interpretive inputs, such as inconsistent supplier formats, free-text emails or terminology that needs interpreting	**→ where AI may add value**
Material judgement, such as whether an estimate is acceptable to report or how to treat a gap	→ human judgement

> **Engineer the predictable parts. Use AI selectively where it adds value.**

2 REDUCE *What disappears*

Stop unnecessary repeat processing, whatever the method.

These apply whether a step is handled by integration, extraction, rules or AI.

- Documents or data unchanged since the last cycle, processed again anyway.
- The same supplier, meter or account mapping worked out again, rather than reusing one already validated.
- The same invoice or spreadsheet arriving by more than one route and processed twice.
- The same data cleaned or reformatted at several stages.
- A whole document set rerun because a few records changed.

3 OPTIMISE *What is handled differently*

Handle the justified AI work in proportion.

- Routine, high-confidence cases handled with lightweight processing.
- Uncertain cases escalated to more capable AI or to human review, rather than every document sent through the heaviest route.
- Context and output limited to the fields actually needed, rather than whole documents processed or summarised.
- At sufficient scale, the overall input-processing workflow may itself warrant redesign, not just lighter individual steps.

The point is the order of the questions. Starting from “where can we add AI?” risks sending every messy input through AI, including inputs that simpler methods would handle and inputs that did not need processing again. Starting from Choose and Reduce shrinks the problem before any AI is sized for it.

> **Takeaway.** AI resource efficiency is not only about lighter individual AI interactions. Sometimes the larger opportunity is to redesign the workflow so that less unnecessary computational activity is needed. The people involved may also gain a simpler workflow, but that is a co-benefit rather than the measure being managed.

Three kinds of evidence

Three different questions

How does an organisation know its management of AI demand is working, and what can it actually see? No single measure answers both questions. This guide distinguishes three kinds of evidence. Each answers a different question and none substitutes for another. They are options to combine, not levels to climb.

	MANAGEMENT ASSURANCE	COMPUTATIONAL INDICATORS	ENVIRONMENTAL RESOURCE INFORMATION
The question it answers	Are the expected ownership, decisions and actions actually happening?	How is observable AI activity changing?	What physical resource use can reasonably be connected to that activity?
Typical evidence	Important uses identified; named owners; objectives or intended outcomes set; actions completed; exceptions and review dates visible; guidance and training that carry resource-efficiency expectations	Calls, tokens or jobs; model or service tier; retries; input and output volumes; credits or cost, with care	Energy, GHG emissions, water, embodied hardware, and location and time effects. No single standard or tool reviewed covers them all.[1]
What it can show	Whether resource efficiency is being actively and competently managed	Whether demand is growing, shifting or apparently wasteful[2]	The closest view available of the physical resources behind AI use
What it cannot show	Any quantified improvement, computational or physical	Environmental impact. The energy behind a call or token varies with model, hardware and utilisation; cost is a commercial term, not a physical proxy	In most service arrangements today, the impact of your specific task (see page 13)
What it depends on	Specific answers from accountable owners	How you use AI: direct model calls give token counts by model; AI embedded in software usually gives prompts, messages or credits[3]	What providers publish, and the choices behind each figure (see page 14)
Weaker vs stronger	"IT has AI covered" or "we have an AI policy" **vs** "these uses matter, this is who owns each, and this is when they are next reviewed"	A single spend or credit total **vs** a few counts tied to an important use, labelled as diagnostics	One published average applied to everything **vs** a figure whose boundary and method you can state

How to use the evidence

Management assurance here means internal control and management review. It is not an external assurance engagement, a certification or a formal audit programme. It can often be gathered soonest, because it depends on people rather than on data a supplier may not provide. AI guidance and training belong here only where they carry resource-efficiency expectations, such as asking whether AI is needed for a task, discouraging duplicated work or reserving the heaviest capabilities for work that needs them. Material limited to permitted use, confidentiality, output checking or human oversight shows that AI use is governed. It is not, by itself, evidence that resource efficiency is being managed. Even where those expectations are present, guidance and training show capability being built. They do not show that computational demand has fallen or that any environmental improvement has followed.

Computational indicators add a view of whether demand is moving the way decisions intended. They stay useful even as provider reporting improves. An organisation's own activity data is continuous and tied to its own decisions, whereas provider figures arrive after the period, aggregated and allocated by methods the buyer does not set. The usage data you need may exist somewhere in the chain but may not be exposed to you, retrievable in practice or detailed enough to act on.

NOTES

1 Water is the most frequent omission among the standards and tools reviewed. Regulation (EU) 2024/1364; Green Software Foundation (2024).

2 A count contains no energy, carbon-intensity or hardware data, and the energy behind a token varies with the model, hardware and set-up. Vellaisamy et al. (2026); Chung et al. (2026).

3 Based on provider documentation reviewed in October 2026. A few embedded products report more detail.

Why one AI task is hard to trace

One visible task, four layers of information

Environmental information is the evidence closest to the question that matters. It is also the hardest to connect to one organisation's use. The reason is structural rather than a matter of effort: a single visible AI task is assembled from several layers, and each is run by someone who holds only part of the information. An everyday example makes this easier to see.

EXAMPLE

A marketing colleague asks the AI assistant in a design platform to **"create a campaign graphic for our spring energy-saving week"**. A few seconds later, several options appear. Behind that single action, the information needed to trace its environmental footprint sits in four layers:

LAYER	WHAT IS KNOWN HERE	WHAT IS TYPICALLY NOT VISIBLE FROM HERE
1 **Purpose** your organisation	Why the graphic was needed, who asked, how often campaigns like this recur, whether the output was used	What happened after the button was pressed
2 **Workflow** the design platform	The immediate task as typed ("create a campaign graphic"), and how it was broken down: interpreting the brief, generating images, producing variations, resizing, perhaps calling other models or services	The wider business purpose and context behind the task, and often the physical detail behind the models it uses
3 **Model or service provider**	How much model activity each call involved, such as text versus image generation or extended reasoning	How those calls map back to your business task, and whether they served a useful outcome
4 **Infrastructure**	Hardware, electricity, cooling and water, location, and the fact that all of it is shared with many other customers	Which part of that activity was your graphic

This is an illustrative sketch, not a description of how any particular product is built.

> **No single party necessarily holds the whole picture.** Tracing the full footprint of one visible task means joining information held by different organisations, and today there is generally no established route for passing it back along the chain to the buyer.[4] In practice, one company may sit in more than one layer, and some requests pass through several further services.

What this means in practice

For a practitioner, the implication is practical rather than technical. A complete footprint for one task is unlikely to be available from any single source. The useful question becomes what the information that does exist can and cannot tell you.

NOTES

4 Where customer allocations exist, they are made within the provider's own cloud. AWS (2025); Google (2026b).

Making sense of environmental numbers

Five questions to ask of any figure

The market is beginning to respond. Providers are publishing new forms of AI environmental information, each designed to answer a particular question. These are valuable early efforts to make previously invisible impacts visible. The practitioner's task is to work out what question a figure was designed to answer before using it for something else.

ASK	WHY IT MATTERS	A REAL EXAMPLE *as checked October 2026; provider reporting changes quickly*
What exactly is being measured?	Energy, carbon and water are different things, and a figure may cover one kind of activity but not another	Google's published per-prompt figures cover energy, carbon and water for a median text prompt in its consumer Gemini app. They do not cover image generation or model training.[5]
What is inside and outside the boundary?	Two honest figures for the same activity can differ widely depending on what they include	Microsoft's emissions dashboard for Microsoft 365 reports emissions by application, but states that Copilot for Microsoft 365 is not currently included. Google's own analysis found a fuller boundary gave more than twice the energy per prompt of a narrower one.[6]
Was it measured, estimated, averaged or allocated?	A figure that can be calculated has not necessarily been measured	AWS gives customers carbon figures by account, service and region, apportioned from shared infrastructure using usage- or revenue-based allocation. AWS itself notes that tracking emissions by individual workload is complex.[7]
Does it represent my actual use?	An average or benchmark describes a typical case, not your workload	Google, OpenAI and Mistral published per-prompt figures in 2025 using different, and in one case undisclosed, methods. Google states its figure does not represent any specific prompt, and an independent comparison concluded the three cannot be meaningfully compared. Long documents, image generation or agent-style tasks may look quite different from a typical prompt.[8]
Is it suitable for the decision or claim I want to make?	A figure good enough for awareness may not be good enough to compare suppliers or support an external claim	A published average may help a team understand the order of magnitude involved. It does not substantiate a claim about the organisation's own AI footprint.

Provider figures are not wrong. They answer the question they were designed to answer, which may not be yours. The same caution applies to ratios. Relating resource use to useful output needs a functional unit, which works for countable services but is not settled for knowledge work, so in our view per-prompt or per-token ratios should not be presented as productivity-adjusted environmental measures for professional work.[9]

PRACTITIONER NOTE

Make the judgement

Questioning a figure is not the end of the job. The practitioner still has to decide three things:

- **Is it fit for the purpose I have in mind?**
- **Does it need a qualification when I use it?**
- **Is more sophisticated measurement worth pursuing?**

More sophisticated measurement is more worthwhile where AI use is significant, where the organisation controls the model call or infrastructure and can act on better data, where an external claim depends on the figure, or where a supplier can provide attributable rather than averaged information. Where these conditions do not apply, management assurance and available computational indicators will usually provide a proportionate basis. Gaps in environmental information are better recorded and raised with suppliers than filled with precise-looking numbers.

NOTES

5 Self-reported and not independently verified. Describes a median text prompt in the consumer Gemini app, May 2025. Elsworth et al. (2025).

6 Stated on Microsoft's own page (checked October 2026). Cloud carbon tools do cover AI sold as a cloud service. Microsoft (2025).

7 AWS allocates emissions by account, service and region. It describes tracking by individual workload as complex. AWS (2025).

8 An independent comparison: the three used different methods and assumptions. Luccioni and Alves da Costa (2025).

9 A functional unit is the unit of useful output that resource use is divided by, e.g. per user or transaction. Green Software Foundation (2024).

Bringing evidence back to management

What management should be able to see

Checking tells an organisation what is happening. The value of that evidence lies in what is done with it: capturing what has been learned, bringing it back to the people accountable in a form that suits the organisation and the significance of its AI use, and using it to make decisions. Management should be able to see what has changed, what is working, what remains unresolved and what needs a decision.

That includes an honest view of the evidence itself: what is measured, what is estimated and what is not yet known. Limits stated clearly are information in their own right. They show where to accept a gap, where to qualify a figure and where to ask a supplier for more.

BRING BACK…	SO THAT MANAGEMENT CAN SEE…
Important uses, and changes in their scale or nature	what has changed, and whether attention and priorities still fit
Progress against the objectives and actions agreed	what is working, and what is not
Exceptions, gaps and unresolved ownership	what needs a decision or escalation
Computational indicators, labelled as diagnostics	where demand is moving, and where it may be wasteful
Environmental information, with its boundary and basis stated	what is known about the physical resources involved
What is not known, and why	where supplier information, expectations or measurement need attention

Internal reporting and external disclosure

The same evidence can serve two purposes. Internal reporting is how the organisation manages performance and makes decisions. External disclosure is how it communicates beyond itself. The two are connected, but they are not interchangeable.[1]

	INTERNAL REPORTING	EXTERNAL DISCLOSURE
Purpose	Manage performance and make decisions	Provide transparency, demonstrate management practice and respond to stakeholder expectations
Audience	Management and governance	External stakeholders; where a reporting regime applies, its defined users, which differ between regimes
Threshold	What management needs in order to act	Materiality, stakeholder relevance and any applicable reporting obligations
Evidence expectation	Enough to support a decision, with limits stated	Appropriate to what is communicated; where impacts are quantified, a clear boundary, method, assumptions and limitations

NOTES

1 Disclosure standards are written for defined users and apply a materiality threshold. IFRS S1 (2023).

Communicating what is being managed

What can be externally disclosed before impacts are fully quantified?

Organisations can communicate how AI resource efficiency is being managed even where its environmental impact cannot yet be quantified reliably. That account can cover:

- how the issue is governed and who is responsible;
- what the organisation is aiming for and what it is doing;
- what progress has been made; and
- where its information is still incomplete.

This kind of account is useful in its own right. It also fits the way sustainability disclosure standards already ask organisations to explain how governance oversees their targets and metrics.[2]

Quantified environmental impacts call for more care. Where they are disclosed, the boundary, methodology, assumptions and limitations should be clear, and the evidence should be appropriate to the claim being made.[3] Incomplete information does not set aside reporting requirements that already apply.

Some organisations may therefore start to investigate how AI-related emissions are reflected in their GHG inventory. As the GHG inventory note explains, an inventory serves an accounting purpose: it shows an organisation's emissions and how they change over time. It is less suited to judging the effect of individual actions, such as how efficiently one AI workflow is run, although that gap may narrow as supplier-specific data improves. The practical task is to know which question a number answers before deciding what to do with it.[4]

PRACTITIONER NOTE

Where might AI appear in a GHG inventory?

Position as at October 2026

CATEGORY. Purchased hosted AI services are, as purchased services, ordinarily accounted for in **Scope 3 Category 1 (purchased goods and services)**. Classification may differ where an arrangement is closer to a leased asset. No AI-specific guidance was identified in the GHG Protocol material reviewed, and Category 1 is within the Protocol's current revision programme.[5]

METHODS. The guidance names four calculation methods: supplier-specific, hybrid, average-data and spend-based. It chooses between them by purpose and data quality rather than a fixed ranking.

Spend-based estimates need particular care

A spend-based figure multiplies what was spent by an average emission factor per unit of spending for a broad economic sector, taken from input-output data for an earlier reference year. AI services are changing much faster than that:

- the price of a given level of AI capability has been falling steeply year on year;
- at least one provider reports large efficiency gains within a single year; and
- AI-specific supplier information remains incomplete.

A spend-based figure moves with price and spend, not with how much computing is done or how efficiently. As prices fall, the figure can fall even while an organisation's computational demand, or the physical resource use behind it, is rising. The reverse also holds: paying more can raise the figure without any change in the work done.

VERDICT. A Scope 3 GHG estimate can represent one environmental impact associated with the resource use behind an AI service, but **it should not be treated as a resource-efficiency metric.**[6]

The question for the next review

Has the scale or nature of AI use materially changed since the last review?

A material change may warrant a stronger response even where efficiency per activity has improved.

NOTES

2 IFRS S1 asks organisations to disclose how governance bodies oversee targets and monitor progress. IFRS S1 (2023), paragraph 27.

3 The UK Green Claims Code expects "robust, credible, relevant and up to date evidence". Competition and Markets Authority (2021).

4 Inventory accounting is not designed to quantify the effect of an organisation's individual actions. GHG Protocol (2023).

5 The Scope 3 Standard is being revised and merged with ISO 14064-1. Consultation expected 2027, publication 2028. GHG Protocol (2011, 2013, 2026).

6 The price of a fixed level of AI performance fell by factors of 9 to 900 over a year. That is price, not energy. Epoch AI (2025); GHG Protocol (2024).

Wider environmental perspective

Efficiency and total demand

That question matters beyond any single review. An organisation can show that each of its important AI uses now runs more efficiently than before, and still find that its total AI use, and the resources behind it, have grown. Per-unit efficiency can improve while absolute resource demand still rises: lower costs, new applications and wider use can add more activity than efficiency removes. That is not inevitable, and efficiency does real work, because without it resource use would be higher still. But the two are different questions, and an answer to one is not an answer to the other.[1]

The approach in this guide judges each use against the outcome it serves. That is the right test for managing demand, but it has a boundary.

Beyond the boundary

Beyond that boundary sit wider questions. The environmental picture is broader than GHG emissions: energy, water, hardware and where infrastructure is located all matter. Whether AI avoids more impact elsewhere than it causes depends on what would otherwise have happened. Judging that, or AI's net environmental effect, needs much broader boundaries than one organisation's management of its own demand.[2] In our view, resource-efficiency management is necessary but not sufficient on its own. These wider questions matter. This guide acknowledges them but does not attempt to resolve them.

Putting it into practice

Organisations do not need complete visibility of AI's environmental impact before they begin managing its resource use, and ESG practitioners do not need to become AI engineers to play their part. A practical way to begin is to pick one significant AI use and take it through the approach once: make it visible enough to manage, give it clear accountability, challenge unnecessary demand, and bring evidence back so the organisation can learn and adapt.

How much management this warrants will differ from one organisation to the next. As better information becomes available, it should strengthen that management rather than hold it back. A management system does not by itself guarantee environmental improvement, but it creates the conditions to recognise demand, challenge waste and respond as use changes.

The first step is making resource efficiency part of how AI is managed — and learning from there.

NOTES

1 Between 2010 and 2018, data-centre computing rose about 550% while energy use rose about 6%. Reviews find that rebound reduces efficiency savings but does not generally cancel them. Masanet et al. (2020); Gillingham, Rapson and Wagner (2016).

2 Standards for avoided-emissions claims require comparison with the most likely alternative and an assessment of rebound. ITU-T L.1480 (2025); WBCSD (2023).

Methodology

How this guide was developed

This guide is evidence-informed practitioner guidance. Compass Works developed it for environmental, sustainability and ESG practitioners whose organisations use AI but do not build it. The organisational context is mainly UK and European, although the management approach applies more widely. It is not intended as exhaustive coverage of the environmental science or engineering of AI.

The guide was developed from a defined publication structure and claim inventory. Substantive propositions were checked against official statistics, legislation, standards, provider documentation, professional guidance and peer-reviewed research. Where the evidence did not support a proposition as first framed, it was qualified, softened or removed.

Evidence and judgement

The guide treats four kinds of statement differently, so that readers can tell what rests on evidence and what is our judgement.

KIND OF STATEMENT	WHAT IT MEANS	HOW IT APPEARS
External evidence	A factual claim supported by appropriate external sources	Stated directly, with a source
Derived synthesis	A conclusion formed by combining verified evidence from more than one source or field	Stated as our reading of the evidence
Professional reasoning	Practitioner interpretation grounded in environmental-management experience	Stated as reasoning, with no source claimed
Compass Works position	What we consider a responsible or proportionate response	Marked “in our view”

Evidence selection and currency

Evidence selected for relevance. Some questions required substantial research but only a short explanation in the guide. Other material was investigated and left out where it did not materially improve practitioner understanding or action. The space a point receives does not indicate how much research sits behind it.

Notes and references. The notes at the end of each page give qualifications and context for the careful reader. The sources they cite, and the other principal documents consulted, are listed in the References.

A fast-moving subject. Provider disclosures, standards, accounting guidance and statistics in this area are changing quickly. Time-sensitive sources were checked in September and October 2026. Figures carry their dates and may need updating in future editions.

Limitations and caveats

This guide sets out a management approach. It does not:

- establish that AI use is environmentally material for every organisation;
- provide an AI emissions factor, or a method for calculating one;
- assume that reducing computational activity creates an immediate, one-for-one physical saving;
- prescribe one ownership, governance or reporting arrangement;
- teach AI engineering, or assign technical execution to ESG;
- rank providers, standards or methodologies; or
- determine AI’s overall net environmental effect, or quantify rebound.

References

A selected list of the documents referred to in preparing this guide. Provider technical documentation from Anthropic, AWS, Google, Microsoft and OpenAI was also reviewed in September 2026. Full list available on request.

AI in organisations and the infrastructure behind it

Department for Energy Security and Net Zero (DESNZ) (2026). *Data centre electricity consumption in Great Britain, 2020 to 2024.* Energy Trends special article.

Elsworth et al. (2025). *Measuring the environmental impact of delivering AI at Google scale.* Google.

European Commission (2024). Commission Delegated Regulation (EU) 2024/1364 on a common Union rating scheme for data centres.

Eurostat (2026). *The use of artificial intelligence (AI) technologies in the European Union: key results.*

Google (2026a). *Gemini AI features now included in Google Workspace subscriptions.* Google Workspace Help.

International Energy Agency (IEA) (2026). *Key Questions on Energy and AI.*

Li, Yang, Islam and Ren (2025). Making AI less "thirsty": uncovering and addressing the secret water footprint of AI models. *Communications of the ACM.*

Microsoft (2026). *Microsoft 365 pricing and packaging updates.*

National Energy System Operator (NESO) (2026). Written evidence DCU0081 to the Environmental Audit Committee inquiry on data centres. UK Parliament.

Office for National Statistics (2026). *Artificial intelligence in UK businesses: 2023 to 2026.*

Variability, visibility and measurement

Amazon Web Services (AWS) (2025). *AWS Customer Carbon Footprint Methodology,* model version 3.0.

Chung et al. (2025). The ML.ENERGY Benchmark: toward automated inference energy measurement and optimization. arXiv.

Chung, Wu, Ma and Chowdhury (2026). Where do the joules go? Diagnosing inference energy consumption. arXiv.

Dauner and Socher (2025). Energy costs of communicating with AI. *Frontiers in Communication,* 10.

Google (2026b). *Carbon Footprint reporting methodology.* Google Cloud.

Google (2026c). *View your organization's carbon footprint report.* Google Workspace Admin Help.

Green Software Foundation (2024). *Software Carbon Intensity (SCI) specification.* ISO/IEC 21031:2024.

ISO/IEC (2026). ISO/IEC 30134-2:2026, *Data centres: key performance indicators, Part 2: Power usage effectiveness (PUE).*

ITU-T (2025). Recommendation L.1450: *Methodologies for the assessment of the environmental impact of the ICT sector.*

Lawrence Berkeley National Laboratory (2026). *United States Data Center Energy Usage Report: 2025 Update.* LBNL-2001758.

Luccioni and Alves da Costa (2025). *What kind of environmental impacts are AI companies disclosing? (And can we compare them?)* Hugging Face.

Luccioni, Jernite and Strubell (2024). Power hungry processing: watts driving the cost of AI deployment? *ACM FAccT '24.*

Microsoft (2025). *Connect to Emissions Impact Dashboard for Microsoft 365.* Microsoft Learn.

Oviedo et al. (2025). Energy use of AI inference, efficiency pathways, and test-time scaling. *Joule* (preprint).

Vellaisamy, Lam, Blanton and Shen (2026). Characterization of request and token energy costs for LLM inference workloads. arXiv.

Management, governance and proportionality

Committee of Sponsoring Organizations of the Treadway Commission (COSO) (2013). *Internal Control: Integrated Framework.*

European Commission (2023). Commission Delegated Regulation (EU) 2023/2772, Annex I, ESRS 1 General Requirements.

European Parliament and Council (2009). Regulation (EC) No 1221/2009 (EMAS), Annex I.

European Parliament and Council (2024). Regulation (EU) 2024/1689 (Artificial Intelligence Act).

Gillespie et al. (2025). *Trust, attitudes and use of artificial intelligence: a global study 2025.* University of Melbourne and KPMG.

Government Digital Service (2025). *Artificial Intelligence Playbook for the UK Government.*

Institute of Internal Auditors (2020). *The IIA's Three Lines Model.*

Institute of Internal Auditors (2024). *Global Internal Audit Standards.*

Institute of Sustainability and Environmental Professionals (ISEP), formerly IEMA (n.d.). *Skills Map and Membership Standards,* v1.0.

International Sustainability Standards Board (2023). *IFRS S1 General Requirements for Disclosure of Sustainability-related Financial Information.*

ISO (2023). ISO/IEC Directives, Part 1, Annex SL, Appendix 2: *Harmonized structure for management system standards.*

ISO (n.d.). *Exploring MSS.* Web page on the harmonized approach to management system standards.

National Institute of Standards and Technology (NIST) (2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0).* NIST AI 100-1.

Office for Product Safety and Standards (2014). *Regulators' Code.* UK Government.

SFIA Foundation (2024). *SFIA 9: Sustainability (SUST).*

The demand hierarchy

European Commission (2021). Commission Recommendation (EU) 2021/1749 on Energy Efficiency First: from principles to practice.

European Parliament and Council (2008). Directive 2008/98/EC on waste, Article 4 (waste hierarchy).

IPCC (2022). Demand, services and social aspects of mitigation. Chapter 5 in *Climate Change 2022: Mitigation of Climate Change.* Working Group III, Sixth Assessment Report.

Accounting, disclosure and wider effects

Competition and Markets Authority (2021). *Making environmental claims on goods and services* (Green Claims Code, CMA146).

Competition and Markets Authority (2026). *Making green claims: getting it right, across the supply chain.*

Epoch AI (2025). *LLM inference prices have fallen rapidly but unequally across tasks.*

GHG Protocol (2011). *Corporate Value Chain (Scope 3) Accounting and Reporting Standard.* WRI and WBCSD.

GHG Protocol (2013). *Technical Guidance for Calculating Scope 3 Emissions,* Chapter 1: Category 1, Purchased goods and services.

GHG Protocol (2023). *Inventory and Project Accounting: A Comparative Review.*

GHG Protocol (2024). *Scope 3 Discussion Paper A.1: Inventory Quality* (working draft).

Gillingham, Rapson and Wagner (2016). The rebound effect and energy efficiency policy. *Review of Environmental Economics and Policy.*

ITU-T (2025). Recommendation L.1480: *Enabling the Net Zero transition: assessing how the use of ICT solutions impacts GHG emissions of other sectors.*

Masanet, Shehabi, Lei, Smith and Koomey (2020). Recalibrating global data center energy-use estimates. *Science,* 367(6481), 984–986.

Science Based Targets initiative (2026). *SBTi Corporate Net-Zero Standard.*

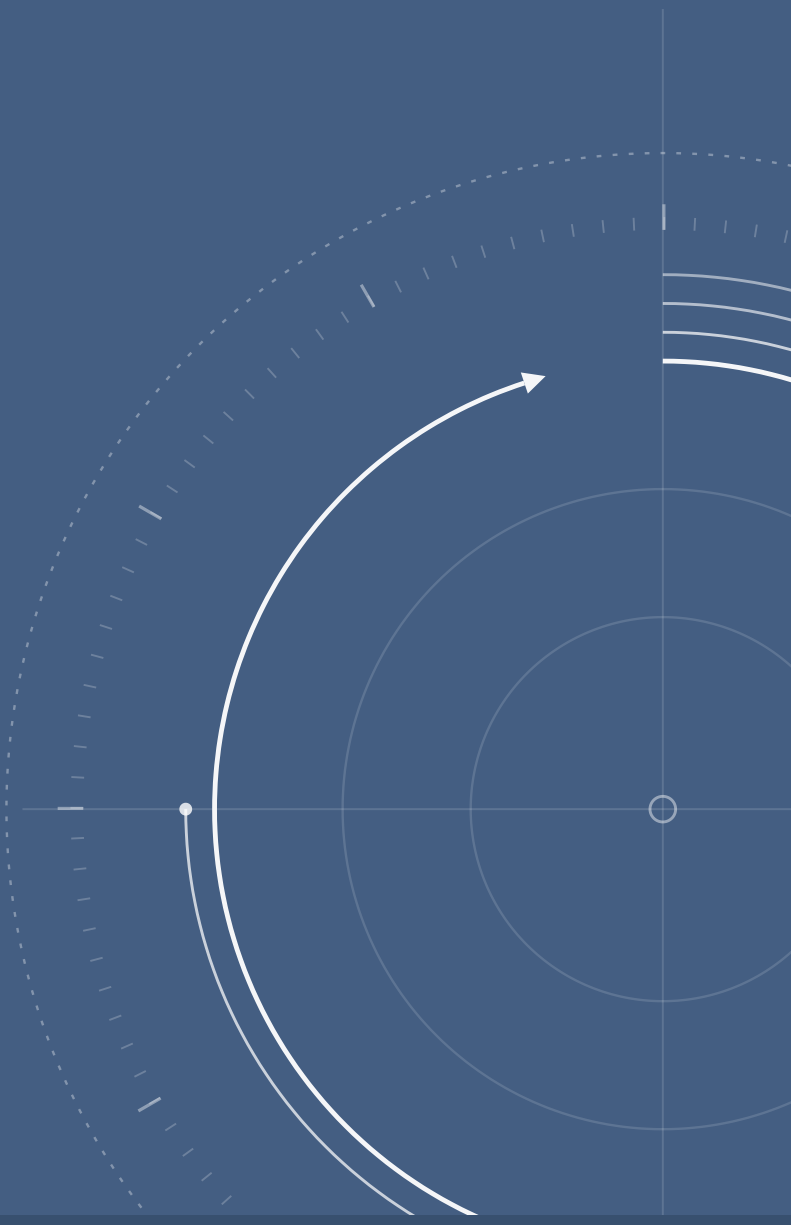
WBCSD (2023). *Guidance on Avoided Emissions.*

COMPASS WORKS

Compass Works is an independent sustainability and data practice working across environmental management, reporting, data workflows and emerging technology.

Through advisory work, original research and practical tools, we help organisations turn complex requirements and fragmented information into clearer decisions, better workflows and systems that make sense.

www.compass-works.com
hello@compass-works.com



PUBLICATION INFORMATION

Authored by:
Jonathan Li, Founder, Compass Works

Published: October 2026

Publisher: Compass Works, Watford, United Kingdom

SUGGESTED CITATION

Compass Works (2026) *Managing AI Resource Efficiency: A practical guide for environmental and sustainability practitioners.* Practitioner Guide [CWI-003]. Watford: Compass Works.

Disclaimer — This publication provides general information and practitioner guidance. It does not constitute legal, accounting or engineering advice, nor does it replace applicable legislation, standards, contractual requirements or organisation-specific assessment. AI technologies, provider disclosures and reporting guidance are evolving and may change after publication. Organisations should assess their own circumstances and seek specialist advice where appropriate.

© 2026 Compass Works. All rights reserved.