Making AI more energy-efficient to ensure a successful transition

By Mathias Abitbol, student at Ecole Normale Supérieure

2nd prize in the 2026 Eco Notepad blog competition

The energy transition hinges on a difficult equation: electrifying the economy whilst decarbonising both production and consumption. AI is making this equation more complex due to the explosion in data centre consumption. But it can also help to solve it, provided two conditions are met: making generative AI more energy-efficient, and deploying specialised AI where it helps reduce carbon emissions. The race to improve AI's energy-efficiency will determine the success of this transition.

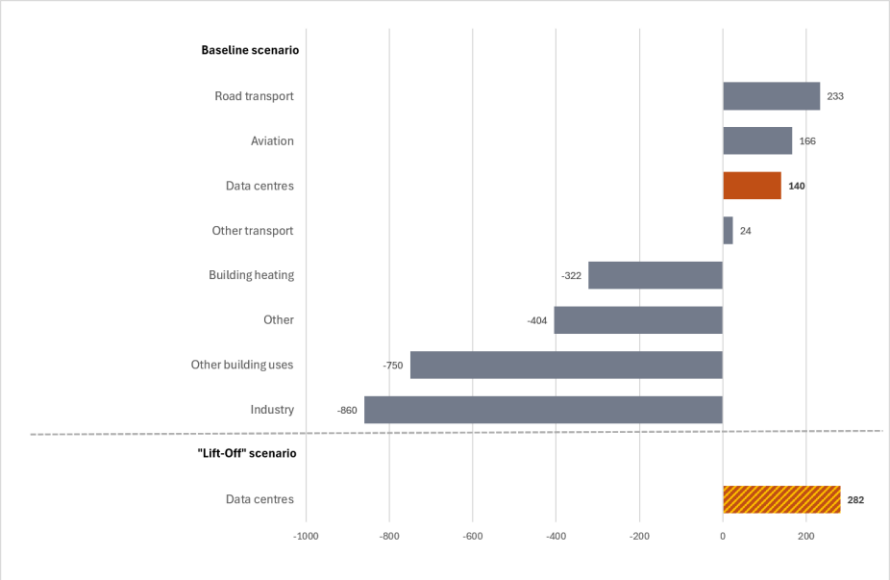
.

AI poses a real threat to the energy transition

The rapid growth of generative artificial intelligence (AI) is driving a new global demand for electricity and water at a time when low-carbon capacity and grid infrastructure are becoming scarce resources. This demand generates negative externalities, by increasing pressure on energy resources and contributing to a rise in greenhouse gas emissions. Data centres will therefore be one of the main sources of rising carbon emissions by 2035.

This pressure is primarily fuelled by the race to improve the performance of generative AI models. Large language models are still at an early stage of development, with the leading players constantly striving to increase the size of these models. This is leading to a sharp rise in energy consumption by data centres. According to the International Energy Agency (IEA, 2025) it was around 415 TWh of electricity in 2024, representing nearly 1.5% of global demand, and could reach around 945 TWh by 2030 in the base case. AI is the main driving force behind this, alongside the more general growth in digital services. Indirect emissions from data centres would thus rise from 180 to 320 Mt of CO_2, an increase of 140 Mt – almost as much as aviation – whilst most other sectors are reducing their emissions. In the 'Lift-Off' scenario, which assumes a faster adoption of AI, the increase would reach 282 Mt, the highest of any sector.

Chart.1 Changes in CO2 emissions (Mt) by sector 2024-30, base and Lift-Off scenarios



Source: AIE (2025)

Note: In the IEA's base scenario, indirect emissions from data centres are projected to increase by 140 Mt of CO_2 between 2024 and 2030, based on assumptions of moderate efficiency gains. In the high-growth scenario (2030, ranging from 180 to 320 Mt in the 'Lift-Off' variant), which assumes accelerated adoption and a faster resolution of energy bottlenecks, the increase reaches 282 Mt.

The surge in consumption is also causing bottlenecks at a local level: data centres are generally connected to grids that are already at full capacity, situated close to city centres. This additional demand increases the carbon intensity of electricity generation, delaying fossil-fuel plant closures and leading to greater reliance on gas and coal for electricity generation, as shown by Bonfiglioli et al.(2025) in their study of consumption trends in US towns and cities hosting data centres.

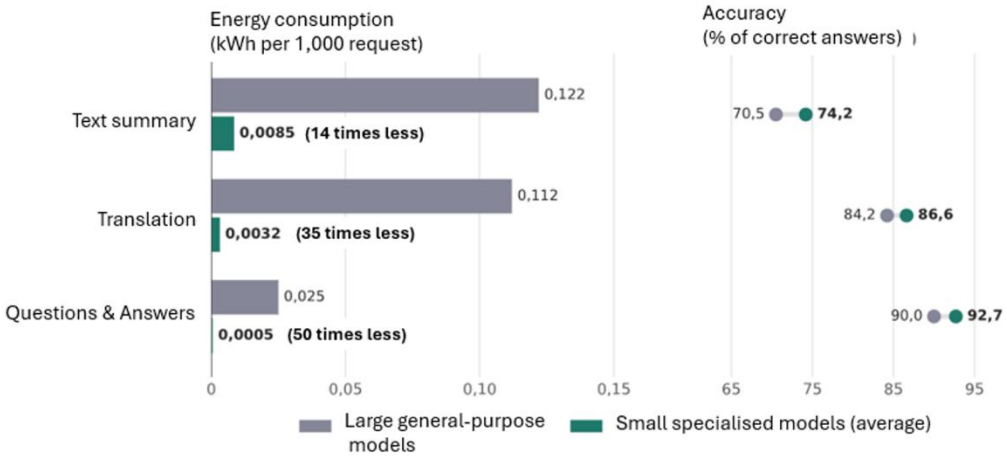
Effective incentives could help green the race to AI

However, these forecasts of strong growth in energy consumption should also be set against the efficiency gains in the digital sector. For every unit of electricity consumed by the servers, data centres consumed an additional 1.5 units for cooling and general operations in 2007, compared with just 0.58 in 2023 (Arcep, 2026).

It therefore appears that the trend observed to date could be reversed by adopting an ambitious institutional framework that promotes energy efficiency and the rational scaling of AI models in line with their intended uses. Initial studies on this topic show that model compression leads to considerable energy savings. According to UNESCO (2025), the adoption of specialised small language models (SLMs) – which consume 14 to 50 times less energy than a large general-purpose model (LLM) – achieves similar or even better results for specific applications, thereby reducing energy consumption by a factor of 10 for tasks such as summarisation, energy consumption by a factor of 10. Furthermore, there is no systematic trade-off between performance and energy consumption: in certain tasks such as translation or question-answering, SLMs perform just as well as specialised models when used on data similar to their training data. There is therefore scope for the development of more energy-efficient models that perform just as well as the most specialised ones, provided that their use is targeted: large general-purpose models would be reserved for tasks

that require their versatility, whilst repetitive and well-defined tasks would be entrusted to more energy-efficient models.

Chart.2. Energy consumption and accuracy depending on the task: a comparison of a large general-purpose model and small specialised models



Source: UNESCO (2025)
Note LLM: Llama 3.1 (8 billion parameters), general-purpose. SLM: task-specific models (61 to 570 million parameters), average values.

The challenge, therefore, is to address this market failure: imperfect information regarding the environmental footprint of the models prevents users from comparing their energy efficiency and from opting for a small, energy-efficient model (SLM) over a large, energy-intensive model (LLM) with equivalent performance. A study analysing several hundred LLM models shows that 84% do not provide any environmental information and only 2% publish a direct measurement of their carbon footprint (Arcep, 2026). Making this negative externality visible to operators and users (through a tax on the energy intensity of inference, efficiency standards or a transparency requirement) would give low-energy models an advantage. SLMs could then replace LLMs for a large proportion of queries.

"Green AI" – a still largely untapped driver of the transition

Viewing AI as a homogeneous technology masks a fundamental distinction: the difference between the impact of generative AI, which receives considerable media attention, and AI specialised in control systems. In the latter case, AI can reduce the energy consumption of existing infrastructures, smooth out peaks in consumption and increase the use of green energy through the coordinated use of vast datasets. These "green" AI systems can generate energy savings and emissions reductions that exceed their own environmental footprint. They therefore offer potential for private gains and collective environmental benefits.

Industry is a first example of great potential. Digital twins and AI-based control systems can adjust and optimise in real time energy-intensive processes in high-energy industries such as ventilation, pumping, heating and pressurisation, etc. According to a model developed by Abitbol et al. (2026),

using econometric analysis, the widespread adoption of AI solutions in those industrial sectors alone where their potential has already been demonstrated could generate economic gains from carbon emission reductions of around USD 940 billion, and up to USD 1,800 billion when energy efficiency gains are taken into account. This latter effect could significantly offset emissions from data centres: according to the IEA (2025), the reduction achieved through the widespread deployment of existing AI applications could be three to four times greater than the emissions from the data centres themselves by 2035.

Table 1 Global well-being gains according to the scope of AI deployment (DICE-2023)

Processes covered	**% Energy CO2**	**Welfare Gain ($bn, PPP)**	
		Climate alone	*Climate + Energy*
Wastewater treatment plants alone	0,4%	75	149
+ industrial drying and galvanising	5%	940	1 800
Total emissions (extrapolation)	100%	18 800	36 000

Source: Abitbol et al. (2026)

Note: If the 6 per cent reduction in measured carbon intensity were to extend to sectors where solutions have proven technological potential (5 per cent of global emissions), the gain in global welfare would amount to $940 billion from avoided emissions alone, and $1,800 billion when the value of the energy saved is included.

Optimising household energy consumption provides a second example. AI is already being used to manage entire areas of household energy expenditure, such as heating and charging electric vehicles. For the latter, the aim is to shift charging to off-peak hours, when electricity is cheaper and produces fewer carbon emissions. In their assessment of such an AI-managed system, Metcalfe et al. (2026) observed that it led to a 42% reduction in household electricity demand during peak hours. AI thus makes flexibility a practical reality: by automating these repeated trade-offs, it shifts the burden to the algorithm and enables a level of efficiency that would otherwise require too much effort on the part of consumers.

In many applications, AI therefore represents an immediate means of decarbonisation, capable of generating significant energy savings at little or no cost. This net gain is, however, subject to two conditions. On the one hand, there is a real risk of a rebound effect: by reducing the energy cost of a service, AI encourages its use, thereby partly offsetting the savings made. On the other hand, there is still uncertainty surrounding the widespread adoption of specialised AI: according to Eurostat (2025), in 2025, just over 5% of European businesses were using AI for optimisation through data analysis. This limited uptake, despite the positive externalities associated with reduced emissions, can be attributed in particular to coordination failures: resistance to challenging long-established professional expertise; market fragmentation, which makes it difficult to connect providers of AI solutions with user organisations; and the low economic value placed on the gains generated (Abitbol et al., 2026).

There are three ways to address these obstacles. First, reward emission reductions: the measured savings could, for example, qualify for energy saving certificates. Second, structure the market by listing solutions with proven gains to connect suppliers and manufacturers. Lastly, assist companies by funding pilot schemes and training, so that AI enhances business expertise rather than threatening it. If these conditions are met, AI will not hinder the transition, and could even speed it up.