

Big AI Models, Bigger Footprints? The *Environmental Impact of SLMs vs. LLMs*

A Comparative Look at the **Environmental Footprint** of Small Language Models (SLMs) and Large Language Models (LLMs).

Small Language Models (SLMs)

Large Language Models (LLMs)

Energy and Water Consumption	<p>Smaller carbon footprint with reduced strain on environmental resources:</p> <ul style="list-style-type: none"> Requires significantly less computational power and smaller datasets for training. Delivers 70–80% energy savings during training and deployment, with <1% performance drop. Consumes less water due to smaller model size and shorter training durations. 	<p>Significant contributor to Carbon and Water Footprints.</p> <ul style="list-style-type: none"> Training a single LLM such as GPT-4 can generate over 502 tons of CO₂ emissions. Consumes as much electricity annually as 5 average American households. Training 429,000 CNNs = 263,000 kWh, equivalent to 24 years of energy use for the average U.S. citizen. If current trends continue unchecked, the AI industry could account for up to 15% of global carbon emissions by 2030.
Training Efficiency	<ul style="list-style-type: none"> Pre-trained models (e.g., Hugging Face Transformers) can be fine-tuned with minimal compute and data requirements. Arcee AI's Spectrum technique reduces training time and compute costs by up to 50%. Arcee AI's DistillKit converts large models into smaller, optimized versions, reducing overall energy consumption during deployment. 	<ul style="list-style-type: none"> Training can take several weeks and requires substantial computational infrastructure. Models like GPT-3 consume approximately 1,287 megawatt-hours of electricity, equivalent to the annual energy use of over 100 U.S. households.
Performance	<p>Smaller models can rival larger ones on key tasks with far fewer resources.</p> <ul style="list-style-type: none"> RoBERTa-base, for example, achieves 90% of RoBERTa-large's performance. Arcee AI's Virtuoso-Medium-v2, a 32B distillation of DeepSeek-V3, excels at chatbots and enterprise analysis. With proper distillation, small models offer strong performance without heavy infrastructure. 	<p>Capable of high performance on a broad range of tasks, yet demand substantially more computational power and energy.</p>
Deployment Impact	<ul style="list-style-type: none"> Small Language Models (SLMs) are ideal for fine-tuning and task-specific use cases, delivering strong performance with minimal resources. Their compact footprint enables deployment on devices without GPUs or with limited RAM—powered by tools like Arcee MergeKit. Perfect for low-resource environments, SLMs support broader accessibility and lower environmental impact. <u>Learn how SLMs on ARM CPUs enable greener, faster AI in our latest article.</u> 	<p>Best suited for general-purpose or large-scale enterprise applications that require broad capabilities, but come with significant environmental cost.</p>

Sources: Small language models (SLMs): A cheaper, greener route into AI, Can artificial intelligence technology reduce carbon emissions? A global perspective, Sustainable AI: The Role of Large and Small Language Models, How Much Energy Does AI Use?, Seven Key Advantages of SLM Over LLM for Businesses.

Ready to build smarter, greener AI? Discover how Small Language Models (SLMs) deliver high performance with lower environmental impact—no heavy infrastructure needed. [Read the full article on SLMs on ARM CPUs.](#)