

NOT EVERY PROMPT IS ENERGY HUNGRY

The boom in artificial intelligence (AI) is leading to increased electricity consumption—that much is clear. On the other hand, AI can reduce energy consumption by helping people adopt a more sustainable lifestyle. It's therefore worth taking a closer look, especially at ChatGPT and other chatbots. After all, not every query consumes the same amount of energy. A study commissioned by the SFOE helps provide a better assessment of electricity consumption.



Stadt Zürich



Menü

Deutsch ▼

Zü-Re

Hallo und willkommen bei Zü-Re – deiner KI mit einer Vorliebe fürs Teilen, Weitergeben und Reparieren in Zürich.

Frag mich alles rund ums gemeinsame Nutzen, das Schenken von Dingen oder wie du kaputten Sachen neues Leben einhauchst. Ich bin zwar schon ziemlich clever, aber noch nicht allwissend – also gib mir etwas Zeit, falls ich mal kurz auf dem Kopf stehe und nachdenke. Ich lerne laufend dazu, damit ich dir immer besser helfen kann.

Worüber möchtest du heute mehr wissen?



Homepage of the Zü-Re sustainability chatbot. Zü-Re was the focus of a case study in which researchers examined the electricity consumption of chatbots based on artificial intelligence. Screenshot: B. Vogel



Thanks to machine-based dialogue systems like ChatGPT, Gemini, Copilot, Grok, or Claude, artificial intelligence is on everyone's lips today. A short query ("prompt") on any possible topic is all it takes, and the text starts flowing. Many responses seem astonishingly plausible and are readily accepted by users. And this is even though they come from a machine—more specifically, from a Large Language Model that has been trained on countless texts to the point where it can formulate meaningful texts on its own. The computational effort required to answer a prompt and the associated electricity consumption remain hidden from most users.

More Than Just the Language Model

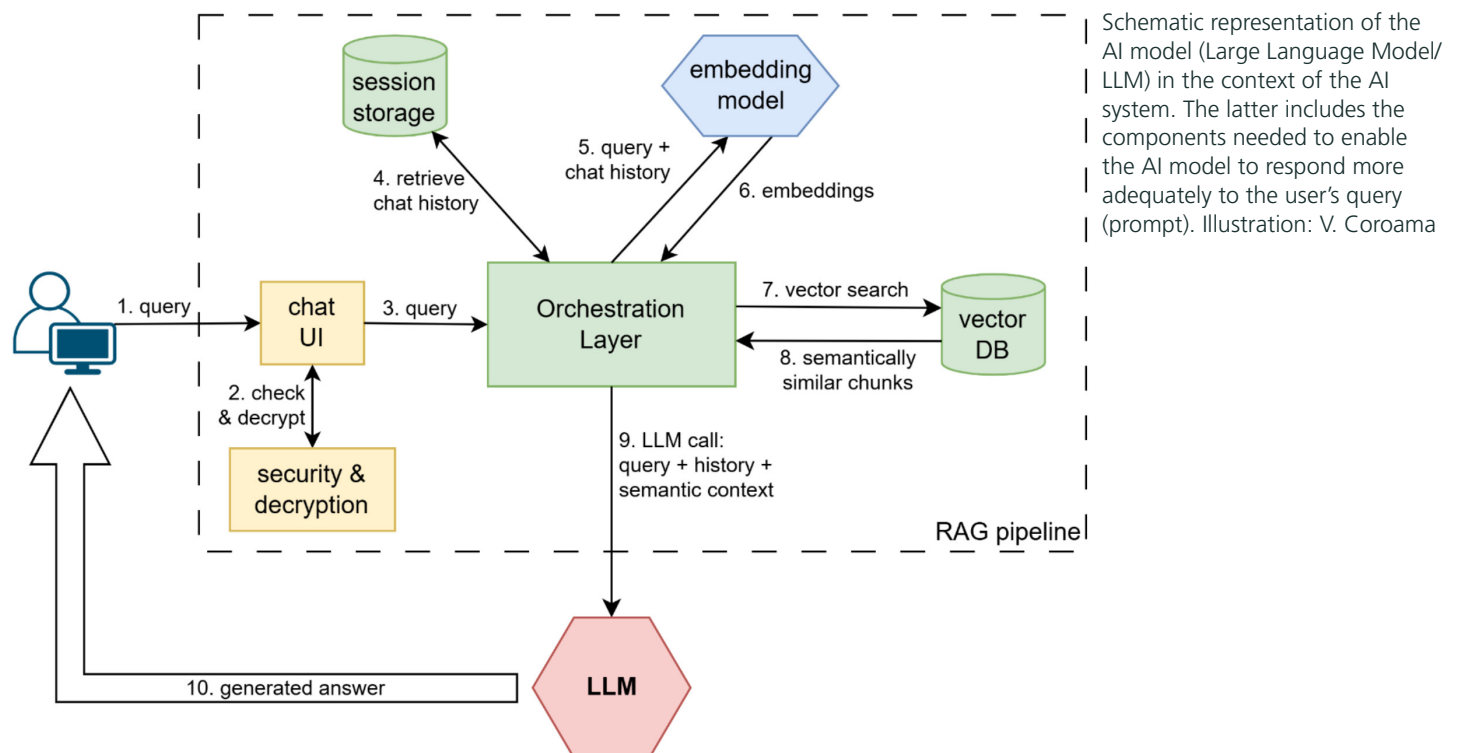
Even for experts, it is not easy to quantify the electricity consumption of chatbots. "When it comes to the Large Language Models on which ChatGPT and similar platforms are based, only the providers know how much electricity it took to train the models and how much electricity is required to respond to a prompt," says Vlad Coroama. A computer scientist by training and the director of the Roegen Center for Sustainability (Zurich), he investigates issues related to the energy consumption of digitalization. On behalf of the SFOE, he has researched various aspects of the electricity consumption of generative AI.

In the study, known by the acronym SAFE-AI, Coroama, to-

gether with Daniel Schien (University of Bristol), developed a methodological framework that helps to better understand and quantify the electricity consumption of AI applications. AI language models consume electricity during training and later when responding to prompts. However, it is not only the AI model that requires electricity, but also the AI system. The AI system, as Coroama understands it, encompasses all IT components necessary for the language model to respond to a user query. These include various internal databases, such as a repository for the content of previous chats that is accessed when new prompts are entered. They also include security and decryption components, as well as a tool that generates words with the same meaning (synonyms) in response to prompts.

Case Study from Zurich

The components of the AI system can have a greater impact on the power consumption of a chatbot query than the consumption of the language model itself. This is illustrated by a case study conducted by the two researchers with support from the City of Zurich: In March 2025, the city launched the sustainability chatbot Zü-Re. It answers questions from the public on the topic of the circular economy: for example, it finds the nearest thrift store for a user or directs the user to a bike shop specializing in e-bikes. As part of the SAFE-AI project, the researchers analyzed approximately 1,100 chats





Data centers have become a topic of public debate due to AI applications. But even setting AI aside, companies—such as those in the financial or pharmaceutical sectors—and research institutions like CERN (photo) continue to require significant computing capacity for data processing. Photo: Shutterstock

that the chatbot responded to in July 2025. Each chat typically consisted of several prompts and the corresponding responses.

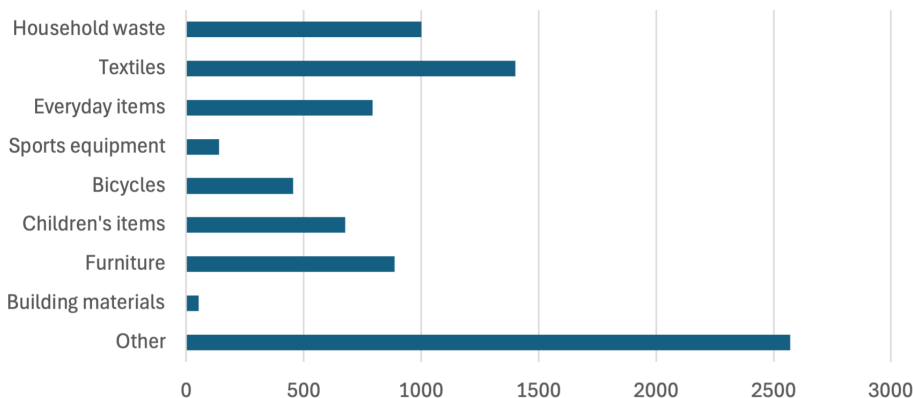
The analysis revealed that 99% of the electricity consumption associated with the 1,107 chats was attributable to the AI system—specifically, the electronic components housed in a colocation data center operated by an external service provider near Zurich that hosts the system. Among other things, the electricity was needed to store documents containing local information about the city of Zurich on servers. The Zü-Re chatbot accesses this data when answering inquiries about local topics: When a user in Zurich enters a prompt, this local information is transmitted along with the prompt to the language model, which is typically hosted by a U.S. data

center today. There, the prompt is processed using the provided contextual information, and the response is then sent to the user in Zurich. Only about 1% of the electricity consumption generated by this chatbot request is attributable to the language model in the U.S. data center. The vast majority is consumed by the IT infrastructure in the Zurich area.

Half a kilowatt-hour per chat

According to the two researchers' calculations, the 1,107 chats in July 2025 consumed a total of 550 kWh of electricity—or about 0.5 kWh per chat. A single chat, consisting of one or more prompts and their corresponding responses, thus requires roughly the same amount of electricity as riding an e-bike for 50 kilometers or watching five hours of television. These figures include the energy consumption of the AI

Topics Covered by the Zü-Re Sustainability Chatbot
(Number of Inquiries)



The inquiries received by the Zü-Re chatbot cover a wide range of topics. Graphic: SAFE-AI/B. Vogel

GROWING ELECTRICITY DEMAND FROM DATA CENTERS

Data centers for artificial intelligence (AI) applications are undergoing major expansion. According to media reports, the four U.S. tech companies Amazon, Meta, Microsoft, and Alphabet—also known as hyperscalers—plan to invest 700 billion U.S. dollars by 2026, primarily in data centers where AI models are trained and operated. High-performance gas-fired power plants, among other sources, are used to meet this electricity demand. Germany, too, is planning a significant number of gas-fired power plants to meet the electricity demands of new data centers.

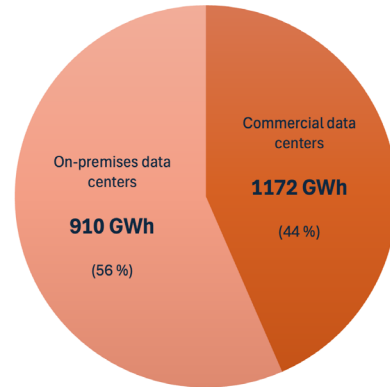
Switzerland has a high density of data centers relative to its population. However, these have not yet been used for training or operating AI models; rather, they are part of the IT infrastructure of conventional companies. A [study](#) commissioned by the SFOE and published in the spring of 2026 estimates the electricity consumption of Swiss data centers for the year 2024 at 2.1 TWh (3.6% of Switzerland's total electricity consumption). Compared to 2019, this represents an increase of 18%. Electricity consumption is expected to continue growing strongly in the future: The study projects an increase of 20 to 50 percent by 2030, and in extreme cases, as much as 70 percent.

Meanwhile, on-premises data centers are becoming less important as companies increasingly outsource their IT. In doing so, they might, for example, use the computing power of cloud providers or operate their own IT hardware in so-called colocation data centers, which do not provide computing power but do offer the necessary data center infrastructure, such as power supply, cooling, fire protection, and monitoring. External data centers are gaining importance and are consuming an ever-larger share of electricity in Switzerland as well. Some of these commercial data centers count large IT companies (hyperscalers) among their customers, companies that do not own their own data centers in Switzerland.

Data center electricity consumption is already quite well documented. A new study by the Roegen Center for Sustainability, also commissioned by the Swiss Federal Office of Energy and titled "Circular Economy Synergies and Trade-offs in Data Centers" (CEST-DC), has now expanded its focus to include the sustainability of data centers. The study addressed issues related to energy efficiency, water and material consumption, and material cycles. It also analyzed the impact of data centers on the power grid.

➤ The final report on the project «Circular Economy Synergies and Trade-offs in Data Centres» (CEST-DC) is available [here](#).

Electricity Consumption of Data Centers in Switzerland, 2024



In a study commissioned by the SFOE, the consulting firm TEP Energy estimated the electricity consumption of Swiss data centers for the year 2024 at 2,082 GWh (or 2.1 TWh). This corresponds to 3.6% of Switzerland's electricity consumption. Graphic: B. Vogel



The CEST-DC study examined various sustainability aspects of data centers. Pictured: Swisscom's data center in Zurich-Herdern, which has been supplying heat to a regional heating network since 2012. Photo: Shutterstock

model (including training), as well as the consumption of the entire AI system (in both data centers and end users' devices, and including the embodied energy required to manufacture these hardware components).

This energy consumption must be weighed against the beneficial effects of the sustainability chatbot. The researchers would have liked to quantify this positive effect but did not receive enough responses to a survey conducted among Zü-Re users. Nevertheless, Vlad Coroama takes a positive view: "0.5 kWh of electricity per chat sounds like a lot. But if someone has their bike repaired instead of buying a new one thanks to the chatbot's advice, it can quickly pay off environmentally. This is especially true for the carbon footprint, since the chatbot uses electricity in Switzerland with a low carbon footprint."

High usage reduces electricity consumption per request

The fact that Zü-Re's electricity consumption—as shown above—is distributed so unevenly between Switzerland and the U.S. does not apply in principle to all chat requests made in Switzerland. The relatively high local electricity consumption of Zü-Re stems from the fact that the chatbot was only sparsely used in the first few months after launch and due to the summer break. During the study period, there were only an average of 36 chats per day. The AI system's electricity consumption, however, is fairly constant and depends only marginally on the number of requests. Since there were only a few inquiries, the electricity consumption per inquiry was very high. If Zü-Re or similar systems are used more extensively in the future, the AI system's electricity consumption per inquiry will decrease. This will also reduce the overall electricity consumption per inquiry for Zü-Re, and the AI model will account for a larger share of that consumption.

Against this backdrop, Vlad Coroama is certain: "To assess the power consumption of chatbots, we need to look at the entire system, not just the language model sitting on a server somewhere in California or Ireland. Local AI systems can account for the lion's share of energy consumption, especially when they are underutilized." Including the AI systems is important, among other things, for corporate sustainability reporting. If companies want to report their environmental footprint, they must also account for the power consumption of the AI system itself, not just that of the AI models being used.

INTERNATIONAL EXCHANGE

Among other things, SAFE-AI and CEST-DC provide methodological foundations for accurately determining the electricity and resource consumption of AI applications and data centers. "The two studies are not only relevant to Switzerland, but they also provide important input to the international research community and international regulatory authorities," says Roland Brüniger, external director of the SFOE's Electricity Technologies research program. Brüniger represents Switzerland in the International Energy Agency's (IEA) Technology Collaboration Program (TCP) 4E. In TCP 4E, experts from 14 countries collaborate to exchange views on issues such as energy consumption and energy efficiency in various application areas and to use the results to support policymakers.

Long prompts consume a lot of energy

In their study, the two researchers develop an approach to quantify the overall energy consumption of chatbot queries. According to their observations, the number of input and output tokens is the most meaningful parameter for estimating energy consumption. Put simply, tokens are the syllables that make up a prompt or the chatbot's response text. The following applies: The energy consumption of the output text increases linearly with the number of tokens, while the energy consumption of the input text increases quadratically (though starting from a much lower value).

For short prompt requests, energy consumption is thus primarily determined by the length of the output text. Summarizing an input text therefore requires less energy than translating it. For longer prompt requests, however, the length of the input text also becomes a significant factor—and at a length of approximately 100,000 tokens, it even becomes the dominant factor. If a user includes, for example, a 300-page PDF in the prompt, power consumption increases significantly. The generation of AI-generated images or videos also requires a comparatively large amount of power because they consist of large output files.

Incidentally, OpenAI and Google estimate power consumption for a typical prompt at 0.3 Wh. That is more than a thousand times less than the 0.5 kWh per Zü-Re chat mentioned above. This is primarily due to the low usage—and thus inefficiency—of the local AI system in the case of Zü-Re. If, for example, the system had been used 100 times

more intensively, it would have consumed the same amount of energy locally, since its capacity was far from being fully utilized. Per chat, energy consumption would thus be 100 times lower, amounting to approximately 5 Wh per chat. Furthermore, the figures from OpenAI and Google refer to a single prompt, not a chat (which typically consists of multiple prompts). Additionally, the consumption figures from OpenAI and Google are defined more narrowly; for example, they do not include the embodied energy required to manufacture the IT components.

- **Final report** on the project "[Sustainability Assessment Framework for the Environmental Impacts of Artificial Intelligence](#)" (SAFE-AI).
- For more **information**, please contact [Roland Brüniger](#), external director of the SFOE's Electricity Technologies research program.
- More **technical articles** on research, pilot, demonstration, and flagship projects in the field of [electrical technologies](#).