Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of the Environment,
Transport, Energy and Communications DETEC

Swiss Federal Office of Energy SFOE

Energy Research and Cleantech

Executive Summary

SAFE-AI – Sustainability Assessment Framework for the Environmental Impacts of Artificial Intelligence

The rapid proliferation of artificial intelligence (AI) brings major opportunities but also growing environmental challenges. Energy, material, and water consumption are increasing, leading in turn to higher greenhouse gas (GHG) emissions and further environmental burdens.

There is thus a growing interest in assessing the footprint of AI as well as its optimisation potentials, especially for companies and organisations deploying AI systems as well as for individuals using it in their daily lives. It is equally important, however, to understand the potential benefits AI might bring beyond its footprint.

The SAFE-AI framework

Dr. Vlad Coroamă (Roegen Centre for Sustainability, Zurich) and Prof. Dr. Dan Schien (University of Bristol) set out to address these questions. Their study, co-funded by the Swiss Federal Office of Energy (SFOE), resulted in an assessment framework called "Sustainability Assessment Framework for the Environmental Impacts of Artificial Intelligence" (SAFE-AI) as well as two validation case studies.

The SAFE-AI framework distinguishes three assessment levels: AI model, AI system, and AI usage. Assessments of AI models have been the focus of the literature so far, because training and all inferences taken together are both energy-heavy processes for frontier model providers. For organisational deployment, however, there is an entire further ecosystem surrounding AI model usage. Its impact can be more important than the footprint of the model itself and is thus necessary for a thorough analysis and reporting. So far, it has not been assessed properly in the literature.

Entire AI system assessment

SAFE-AI thus proposes an AI system assessment workflow for organisations with the following characteristics:

i) It allows for a variety of AI system architectures, including both organisation-internally deployed models as well as externally employed models (e.g., from a frontier model provider such as OpenAI, Anthropic, or Google).
ii) It covers energy consumption, GHG emissions, and water consumption.
iii) In true life cycle assessment (LCA) manner, it allocates to systems in use not only shares of model training and device production, but also of basic research and the development of intermediate, unreleased models by the model providers.
iv) Finally, it provides meaningful and workable defaults grounded in the existing body of knowledge for the impact sources that are not visible to the organisation performing the assessment.

Assessment of individual usages

A further important question is how AI usage can be measured in a meaningful way. The study shows that the number of output and input tokens is particularly suitable for estimating energy consumption. The energy of the widely deployed Transformer models varies linearly with the number of output tokens and with the square of input tokens; however, starting from a very low value for input tokens.

For short queries, energy consumption is therefore mainly determined by the length of the output. This has relevant implications for users. Summarising an input text, for example, thus generally consumes less energy than rewriting it. For longer queries, on the other hand (as they are typical for retrieval-augment generation, RAG, for instance), the length of the input text also plays a role – and from a length of approximately 100,000 tokens on, it even becomes the decisive factor. If, for example, a user includes a 300-page PDF in their input, electricity consumption rises sharply. A comparatively large amount of electricity is also required by AI to generate images or videos.

Typical for an organisation employing external AI models is to receive for each billing period the number of queries as well as the total amount of input and output tokens used in that period. Using average values to derive overall energy usage would be misleading due to the partly non-linear dependencies. SAFE-AI thus presents a stochastic analysis of the token distribution per request, which enables a more robust aggregation at system level. The results of this analysis are included in the workflow (see Chapter 6 of the final report; the more curious practitioners or interested researchers can read the explanatory details in Chapter 5).

Validation via the Zü-Re chatbot

The SAFE-AI framework was validated via “Zü-Re”, a sustainability chatbot developed by the city of Zurich to promote circular economy processes and answer citizens' questions on the topic. The chatbot is a RAG-based AI system consisting of several services running in a local colocation data centre, and which employs two external AI models by OpenAI, presumably from the US.

Its energy, GHG, and water footprints were assessed via the SAFE-AI workflow. The results show a relatively high energy consumption of about 0.5 kWh per chat, which stands in stark contrast to the numbers published by frontier model providers such as OpenAI or Google, which are 3 orders of magnitude lower, around 0.3 Wh / query. As a chat typically consists of several queries, the two do not measure the same. This, however, is not the most important reason behind this difference.

The main explanation for this discrepancy lies in the system’s high local energy consumption. The chatbot was used relatively sparsely in the first few months after it went live, not being yet widely known when the assessment took place. During this period under review, there were only an average of 36 chats per day. A minimum of system components needed nevertheless to run continuously, consuming energy 24/7, thus leading to the high per-usage local energy consumption.

In this context, about 99% of the system’s energy consumption occurred locally, explaining two orders of magnitude as compared to assessments that only take the model’s energy consumption into account. This clearly highlights the importance of assessing entire systems as required by SAFE-AI, which is an issue largely ignored so far. It also shows the importance of the effects of scale in achieving system-wide efficiency: Had the system been used 100 times more intensely, its overall energy consumption would have largely been the same, and the per-usage energy consequently two orders of magnitude lower.

The third order of magnitude is explained by two sources: As mentioned above, the Zü-Re assessment relates the energy consumption to entire chats, which typically consist of several queries (and are the focus of other assessments). But it is also explained by the wider, LCA-compliant system boundaries used: SAFE-AI does not only consider the operational impact in the data centre during training and inference. It also considers the production phase of the microelectronics, the impact of end-user devices as well as the impact of further overhead such as basic research and the development of intermediate, unreleased models.

Indirect effects

Beyond its direct footprint, AI can also induce beneficial effects in other sectors (e.g., through efficiency gains, dematerialisation, or fostering circular economy processes) but also detrimental ones (through, for example, more efficient oil & gas drilling or AI-enabled autonomous vehicles substituting public transportation). These are generally known as indirect effects. They are highly contextual and almost impossible to generalise; their assessment is thus not part of the SAFE-AI framework or workflow.

Instead, the study assesses indirect environmental effects in two archetypal case studies: A clinical AI conversational agent used by the National Health Service (NHS) in the UK, and the same Zü-Re sustainability chatbot of the city of Zurich.

Given the clear procedures of the cataract surgery pathway, the former is a rare example for when the counterfactual scenario (i.e., “what would have happened if it were not for AI?”) is well defined. Its net environmental effect (i.e., direct footprint + indirect effects) could thus be established with a high degree of certainty. The conversational agent reduced face-to-face pre-surgery appointments by 3% and post-surgery ones by a substantial 26%. The reduced patient trips to the four hospitals considered for the assessment saved about 260 kg CO_2 emissions per 100 surgeries.

The second case study is the more typical example when the counterfactual is inherently hypothetical and thus ontologically uncertain. The assessment of its indirect benefits was thus performed scenario-based and compared to the chatbot’s direct footprint. Over one month, the chatbot’s roughly 1100 chats were responsible for about 540 kWh energy consumption, 35 kg CO_2 emissions, and 1,500 litres of water consumption; all of them mainly in Switzerland. The relatively low GHG impact is due to the very clean Swiss power mix, in which low-carbon sources such as hydro and nuclear dominate.

To generate indirect impact scenarios, the topics of the chats were analysed: Numerous referred to textiles and how they might be repaired, enough other chats to bicycles. LCA databases inform on the environmental impacts as well as the typical lifespans of producing textiles and bicycles. It could thus be established that if among the 235 monthly chats on textiles and 90 monthly chats on bicycles, the chatbots recommendations have led to one sweater and one bicycle being worn and used for one year more each, respectively, this would roughly offset the bots’ direct emissions.

The final report is available at https://www.aramis.admin.ch/Texte/?ProjectID=55461.