

Digital

ETH zieht Bilanz nach Start – aber Zukunft ist ungewiss: 420'000 Downloads für Schweizer KI-Modell Apertus

Ein Monat nach dem Launch von Apertus zieht EPFL-Professor Martin Jaggi im Podcast Prompt Zero Bilanz. Die Kritik am ersten Schweizer Sprachmodell nimmt er gelassen und erklärt, warum das Schweizer LLM keine ChatGPT-Alternative ist. Wie gehts jetzt weiter?

Tobias Bolzern

Darum gehts

Als Apertus am 2. September lanciert wurde, war die Euphorie gross: Ein Schweizer KI-Modell, trainiert auf über 1000 Sprachen, transparent, offen und mit legalen Daten. Doch schnell kam Ernüchterung auf: Wer nach Schweizer Bundesräten fragte, erhielt frei erfundene Namen. Das Problem: in der öffentlichen Wahrnehmung wurde Apertus oft mit kommerziellen Produkten wie ChatGPT verglichen – ein unfairer Vergleich.

«Das war ein bisschen enttäuschend für uns», gibt Martin Jaggi. Der EPFL-Professor leitete das Projekt zusammen mit Kollegen aus Zürich und Lausanne. Im Gespräch mit Blick aber hält er dagegen: «Auch anderen offene KI-Modelle wie Lama oder Qwen liegen bei den Schweizer Bundesräten falsch.»

Das Chrüsimüsi zum Start war programmiert. «Wir haben immer gesagt: Wir sind keine ChatGPT-Alternative», sagt Jaggi. «Apertus ist kein fertiges Auto, es ist nur der Motor davon.» Ein Basismodell, ohne Schnickschnack wie Suchmaschinenanbindung. Ganz im Gegensatz zu ChatGPT, dass ein kommerzielles Produkt ist und zudem mit einem Vielfachen an Budget trainiert wurde. Vergleicht man Apertus in der gleichen Gewichtsklasse der Modelle von Meta oder OpenAI, sieht es anders aus: Da liegen die Schweizer Modelle auf Augenhöhe oder haben die Nase gar vorne.

Die ersten Nutzungszahlen? Durchaus beeindruckend. Über 420'000 Downloads verzeichneten die beiden Modelle auf Huggingface.co. Zum Vergleich: Das neuste Qwen-Modell aus China erreichte im selben Zeitraum 4,2 Millionen Downloads: David gegen Goliath.

Bei 15 Hackathons in Zürich, Lausanne und Basel tüfteln Entwickler die vergangenen Wochen an ersten Anwendungen mit dem Schweizer KI-Modell. Von 144 Prototypen laufen 50 mit Apertus im Hintergrund. Hier nur drei Ideen, die an den Swiss AI Weeks entstanden sind:

Die Zukunft ist ungewiss bis düster. Während US-Konzerne Rechenzentren mit der Leistung von zehn Atomkraftwerken hochziehen, wurde Apertus mit einem Faktor 1000 weniger in Lugano trainiert. Zudem hatte die Schweiz Glück, dass sie die nötigen GPUs von Nvidia noch vor dem GenAI-Boom relativ günstig kaufen konnte.

«Längerfristig ist schon die Frage, ob die Schweiz den Aufwand stemmen kann», sagt Jaggi. Der Plan: jährlich ein verbessertes Modell, eventuell in Kooperation mit Partnern in Europa. Das hängt aber auch von den finanziellen Mitteln ab.

Das Vermächtnis soll laut Jaggi bleiben: «Ich hoffe, dass wir mehr Transparenz erreichen können in KI-Modellen.»