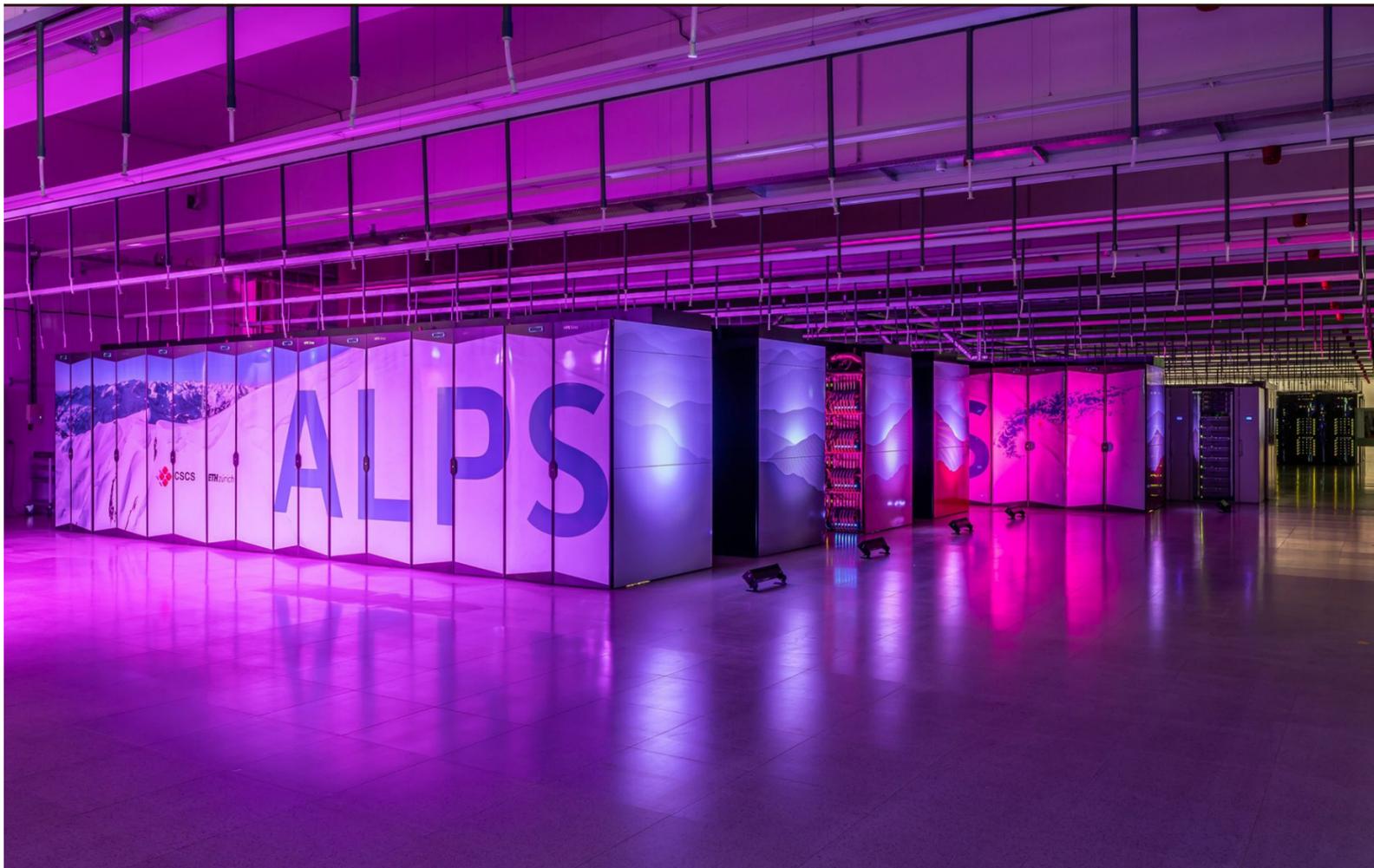


# Kultur, Gesellschaft & Wissen



Hier wurde das Schweizer Chat-GPT trainiert: Supercomputer Alps am Swiss National Supercomputing Centre (CSCS) in Lugano. Foto: Marco Abram (PD)

## Was kann das Chat-GPT der ETH? «Mehr als 1000 Sprachen»

**Neues KI-Modell** Schweizer Hochschulen haben einen Chatbot entwickelt. Wie schneidet er im Vergleich zu Chat-GPT ab? Martin Jaggi, Professor an der EPFL Lausanne, erklärt den «fundamentalen Unterschied».

**Joachim Laukenmann**

Die beiden ETH in Zürich und Lausanne haben einen grossen Chatbot entwickelt. Er soll eine breite Nutzung haben und Innovationen in Wissenschaft, Gesellschaft und Industrie ermöglichen. Trainiert wurde das Sprachmodell auf dem Supercomputer Alps am nationalen Supercomputer-Zentrum CSCS in Lugano. Mit Imanol Schlag von der ETH Zürich und Antoine Bosselut von der ETH Lausanne (EPFL) hat Martin Jaggi das grosse Schweizer Sprachmodell massgeblich mitentwickelt. Er ist ausserordentlicher Professor an der EPFL und Mitglied des EPFL AI Center.

**Herr Jaggi, hat das neue Schweizer Sprachmodell schon einen Namen?**

Ja. Aber wir verraten ihn erst, wenn wir es im Spätsommer öffentlich präsentieren.

**Wie schneidet dieser Chatbot im Vergleich zu Chat-GPT ab?**

Der fundamentale Unterschied ist, dass unser Modell offen ist und Chat-GPT hinter verschlossenen Türen läuft.

**Was bedeutet das?**

Bei Chat-GPT und ähnlichen kommerziellen Angeboten wissen wir nicht, mit welchen Daten sie trainiert wurden, und auch nicht, wie gross die Modelle sind, also wie viele Parameter sie besitzen.

**Wofür stehen diese Parameter?** Vereinfacht gesagt sind das die Verbindungen zwischen den

Neuronen des Modells. Diese werden beim Training des Sprachmodells justiert. Die kleinere Variante von unserem Modell hat 8 Milliarden Parameter und kann damit zum Beispiel auch auf Laptops verwendet werden. Die grosse Variante besitzt 70 Milliarden Parameter. Unser Modell kann man herunterladen, untersuchen und weiterentwickeln. Wir machen auch die Trainingsdaten zu hundert Prozent transparent.

**Was bringt diese Transparenz?** Sie schafft Vertrauen, zum Beispiel für Hochrisikoanwendungen. Da will man wissen, ob ein Chatbot Informationen verzerrt wiedergibt und ob er mit Vorurteilen behaftet ist.

**Verzerrungen und Vorurteile gibt es bei Ihrem Modell nicht?**

Doch. Aber wir können sie transparent machen und besser verstehen. Und dann können wir damit umgehen.

**Was sind denn typische Hochrisikoanwendungen, für die das neue Schweizer Sprachmodell prädestiniert ist?**

Wenn zum Beispiel das Bundesgericht einen Chatbot benutzt, um Informationen in Gerichtsakten zu suchen, ist es wichtig, dass man dem Chatbot vertrauen kann. Ein anderer Anwendungsfall ist ein Start-up, das verhindern will, dass wichtige Firmendaten durch die Nutzung eines Chatbots zu den Big-Tech-Unternehmen in den USA oder in China abfliessen. Oder im medizinischen Bereich: Da ist es heute nicht erlaubt, dass Patientenda-

ten die Schweiz verlassen. Spitäler könnten auf unserem System basierende Chatbots selbst verwalten und nutzen, ohne dass Daten das Spital verlassen. Auch für viele wissenschaftliche Fragen ist ein offenes Sprachmodell essenziell, etwa, um besser zu verstehen, warum Chatbots halluzinieren, also Dinge erfinden, und wie sich das beheben lässt.

**Kann das Schweizer Chat-GPT auch von der Allgemeinheit genutzt werden?**

Ja. Es ist ja das Faszinierende an den grossen Sprachmodellen, dass sie sich sehr breit anwenden lassen. Viele Leute nutzen diese Modelle heute schon privat und in ihrem Job. In diesem Sinn



**Martin Jaggi**  
EPFL AI Center

unterscheidet sich unser Modell nicht von Systemen wie Chat-GPT oder Gemini. Es unterscheidet sich aber in anderer Hinsicht: in seiner Mehrsprachigkeit.

**Auch Chat-GPT beherrscht viele Sprachen.**

Das schon, aber es ist prioritärer für den amerikanischen und chinesischen Markt gemacht und fokussiert daher vorwiegend auf Englisch und Chinesisch. Ich finde es extrem wichtig, dass es ein Sprachmodell gibt, das möglichst alle Bevölkerungsgruppen repräsentiert. Daher haben wir unser Modell so entwickelt, dass

es auf mehr als 1000 Sprachen funktioniert. Darin sind viele seltene afrikanische Sprachen genauso enthalten wie Rätoromanisch oder Schweizerdeutsch. Ein solches mehrsprachiges Modell funktioniert also wie eine Art Interface zwischen all diesen Sprachen.

**Die USA investieren im Rahmen ihres Projekts Stargate 500 Milliarden Dollar in künstliche Intelligenz.**

**Wie gross ist das Schweizer Chatbot-Projekt im Vergleich?** Rein von den Investitionen her können wir mit Stargate natürlich nicht mithalten. Die etwas mehr als drei Monate Training auf dem Supercomputer Alps am CSCS haben rund sieben Millionen Rechenstunden auf Grafikkarten verwendet, was auch etwa den Kosten in Schweizer Franken entspricht. Die kommerziellen Sprachmodelle wie Chat-GPT oder Gemini sind also viel grösser und wurden wahrscheinlich auch mit deutlich mehr Daten gefüttert.

**Diese Modelle wissen also mehr?** Ja.

**Ist das nicht ein Nachteil eines offenen Modells, wenn es beim Einsatz in Forschung und Entwicklung auf weniger Wissen zurückgreifen kann?**

Einerseits ist es für uns sehr ermutigend, zu sehen, dass die Kluft zwischen grossen kommerziellen Modellen und den offenen Modellen bei der Leistungsfähigkeit gar nicht so gross geworden ist wie befürchtet. Aber

ja, es gibt sicher Bereiche, wo wir nicht ganz mithalten können, etwa beim Einsatz des Sprachmodells als Assistent fürs Programmieren. Ein Vorteil unseres Modells ist hingegen, dass es viel effizienter, schneller und günstiger ist. Bei einer wirklich schwierigen und komplexen Frage ist es vielleicht besser, auf ein grösseres Modell zurückzugreifen. Dafür wird dann mehr Strom und mehr Rechenzeit benötigt.

**Wie viel Strom hat das Schweizer Sprachmodell für das Training benötigt?**

Der Strombedarf beim Training war weniger als der Bedarf einer Lokomotive der SBB – Tag und Nacht über einen Zeitraum von rund drei Monaten, gedeckt mit Wasserkraft. Bei der Nutzung des Chatbots ist der Strombedarf allerdings um einige Grössenordnungen geringer.

**Was die Effizienz anbelangt, hat letztlich das chinesische Sprachmodell Deepseek für Schlagzeilen gesorgt, weil es für das Training mit viel weniger Rechenleistung auskam als Chat-GPT oder Gemini. Wie effizient ist das Schweizer Sprachmodell im Vergleich zu Deepseek?**

Wir haben unser Modell in etwa gleich lang trainiert, wie Deepseek trainiert wurde. Im Spätsommer, wenn wir unser Modell publik machen, werden wir sehen, ob es bei der Performance auch vergleichbar ist, wobei unser Fokus sicher stärker auf der Mehrsprachigkeit lag. Exakt vergleichen lassen sich die Modelle daher nicht.

### Männer, ihr nervt mit eurer ständigen Geräuschemission!

**Glosse** Grunzen, Schnauben, mit Fingern trommeln: Auffallend viele Männer geben im öffentlichen Raum laute Geräusche von sich. Bitte aufhören.

Es ist Rushhour im ÖV, aber ich habe den Sechser im Pendler-Lotto: ein freies Viererabteil in der S-Bahn! Doch das Glück währt nicht lange.

Ein Mann setzt sich zu mir – zerstört die Idylle, indem er sogleich ein Geräuschkonzert anstimmte: eine Kakophonie aus herzhaftem Grunzen, Husten, Schnauben, Brummen und Kieferknacken. Verzweifelt will ich mich in die Stille meiner Noise-Cancelling-Kopfhörer flüchten. Aber Mist, kein Akku.

Also bin ich dieser Geräuschkulisse ausgeliefert. Und zunehmend irritiert. Nicht nur, weil es erstaunlich ist, was für einen Lärm ein einzelner Mensch erzeugen kann. Sondern weil dieser Mann diese Körperklänge fast genussvoll von sich gibt. Ohne Rücksicht auf seine Mitmenschen.

Es ist eine Unart, die ich schon seit einiger Zeit beobachte. Die Selbstverständlichkeit vieler Männer im öffentlichen Raum – insbesondere im ÖV –, laut mit den Fingern zu trommeln, mit der Zunge zu schnalzen und die Nase hochzuziehen.

Ähnliche Berichte von Freundinnen bestätigen mir: Falls dieses Phänomen noch keinen Namen hat, ist es höchste Zeit, einen dafür zu erfinden. Denn es nervt – gewaltig.

Vielleicht wäre «akustisches Manspreading» treffend – angelehnt an das bekannte physische Verhalten, bei dem Männer mit breit gespreizten Beinen mehr Platz einnehmen, als nötig oder angenehm wäre. Auf ähnliche Weise beanspruchen manche Männer auch akustisch Raum, oft weit über das hinaus, was viele Frauen sich trauen würden.

Natürlich hat man manchmal einen hartnäckigen Frosch im Hals oder ist erkältet. Aber die Ausrede, dass es sich um eine unbewusste Gewohnheit handelt, zieht hier nicht. Es ist eine Frage des sozialen Miteinanders – und dass man im öffentlichen Raum Rücksicht nimmt. Haben diese lauten Männer schlichtweg ein mangelndes Gespür für ihre Mitmenschen?

Im Falle des grunzenden Mannes im Viererabteil beschleicht mich aber rasch eine andere Vermutung: dass er mit seinen Lauten seine Präsenz betonen – oder cool wirken will. Aus meinem Augenwinkel sehe ich, dass der Herr nach jedem Grunzen meinen Blick sucht. Ich starre auf mein Handy, während er weiterhin sein ganzes Repertoire an auditiver Dominanz auspackt.

Jedes kehlige Räuspern gleicht einem «Hier bin ich! Höre mich!» – Signallaute, die vielleicht in der Tierwelt funktionieren. Im öffentlichen Raum aber die Nerven der Mitmenschen strapazieren.

**Lisa Füllemann**