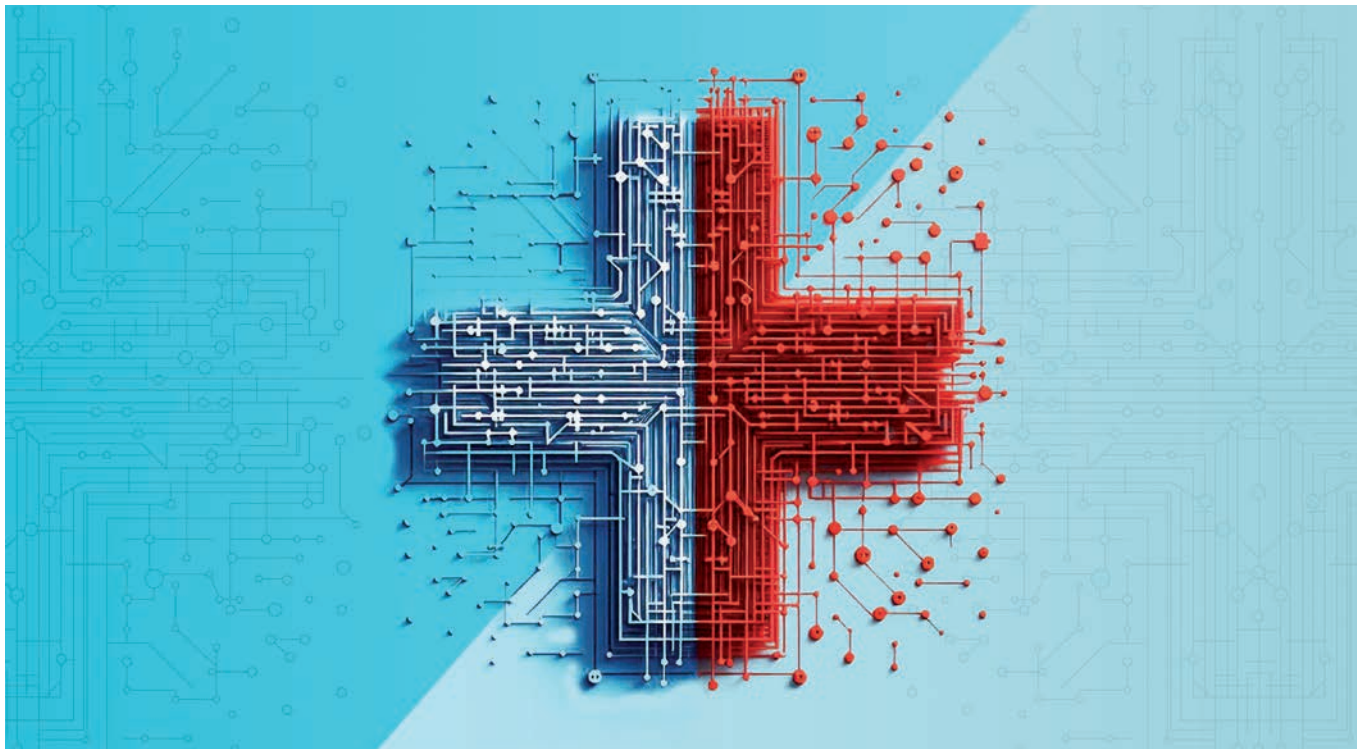


Apertus – ein offenes, transparentes, mehrsprachiges KI-Sprachmodell aus der Schweiz

Stephan Holländer



„ETH_Apertus_Keyvisual_Final-LY2_16-9-ratio-keyvisual.“ Credits: EPFL, ETH Zurich, CSCS / molinari design

› In den Schweizer Bibliotheken ist die verbreitete Verwendung künstlicher Intelligenz (KI) im praktischen Einsatz gering, vor allem im Bereich der Indexierung. Das Thema ist zwar in aller Munde, aber im Programm von Fachtagungen und in Fachzeitschriften noch nicht sehr verbreitet. Wer sich umhört, erhält vielfach die Antwort, dass man den Einsatz von KI in Bezug auf praxisbezogene Anwendungen noch nicht in Erwägung gezogen hat. Für eine kleine Minderheit von Bibliotheksmitarbeitenden ist die Idee in der Umsetzung noch nicht reif genug, um sie weiterzuverfolgen. Einige wenige Bibliotheken planen erste Tests, einige wenige führen Pilotversuche durch, sehr wenige Bibliotheken wenden KI-Lösungen in größerem Maßstab an. Man scheint zu warten, was SLSP als nationaler Bibliotheksverbund als KI-Anwendung anzubieten gedenkt. Eine Arbeitsgruppe mit Mitgliedern aus Bibliotheken, die swisscovery verwenden, hat sich an die Arbeit gemacht.

Sehr wenige Bibliotheken haben versucht, Datenmodelle selbst zu trainieren, um sie in KI-Anwendungen zu testen oder bereits praktisch einzusetzen. Diesen Projekten werden aus mehreren Gründen enge Grenzen gesetzt: Der

Datendurchsatz auf den Servern, die von Bibliotheken benutzt werden, ist begrenzt. Der Zugriff auf Daten ist aus urheberrechtlichen Gründen auf die elektronischen Bestände der eigenen Bibliothek oder den Verbund beschränkt. Nicht zuletzt sind die für diese Projekte eingesetzten finanziellen Mittel beschränkt und die entsprechenden Fachleute müssen gefunden werden.

Trotzdem hat sich seit dem Spätsommer in der Schweiz etwas geändert. Forscher der beiden Eidgenössisch Technischen Hochschulen (ETH) in Lausanne und Zürich haben ein großes Schweizer Sprachmodell (Large Language Model, LLM) mit Namen «Apertus» entwickelt und vorgestellt. Wer sich in der Presse die steckbriefartigen Eigenschaften dieses Datenmodells anschaut, merkt, dass der Name «Apertus» (Offen) Programm ist.

Was ist Apertus?

Apertus ist das erste offene Schweizer Große Sprachmodell (LLM), entwickelt von der ETH Zürich, der EPFL-Lausanne und dem Supercomputerzentrum CSCS (Centro svizzero di calcolo scientifico) in Lugano. Es soll eine offene, transparente Sprachmodell-Alternative zu US- und

China-Modellen bieten. Apertus ist ein reines Textmodell. Es beherrscht mehr als 1.000 Sprachen, darunter auch Schweizerdeutsch und Rätoromanisch, was ein Alleinstellungsmerkmal dieses Modells für die vielsprachige Schweiz ist.

Wie umfangreich ist Apertus?

Apertus existiert in zwei Versionen: Eine im Umfang reduzierte Version mit 8 Milliarden Parametern und eine umfangreichere Version mit 70 Milliarden Parametern. Das größere Modell wurde mit 15 Billionen Wörtern («Tokens») trainiert, was in etwa vergleichbar ist mit der Llama-3-Familie von Meta. Das Training lief während rund drei Monaten auf über 4.000 Grafikkarten des Supercomputers Alps in Lugano.

Welche Funktionalitäten bietet Apertus?

Apertus ist in seiner ersten Version ein reines Textmodell mit Text-Input und Text-Output-Funktionen. Bild- und Ton-Generierung sollen künftigen Versionen vorbehalten sein. Für das Training des Modells konnten 15 Milliarden Textbausteine aus dem FineWeb-Datensatz von Hugging Face sowie aus dem Common Crawl Datensatz genutzt werden. FineWeb-Datensatz stellt hochwertige Vorab-Trainingsdaten für über 1.000 Sprachen bereit. Der Common Crawl-Datensatz enthält urheberrechtlich geschützte Werke und wird in den USA unter Berufung auf das Recht der fairen Nutzung verbreitet. Um mit dem geltenden Urheberrecht in ihren Rechtsordnungen in Einklang zu sein, haben Forscher in Ländern außerhalb der USA mit Verweis auf das Nutzungsrecht des Common-Crawl-Datensatzes Techniken wie das Mischen von Sätzen und Text-/Wortbausteinen angewandt. Diese Textbausteine wurden gefiltert und dem thematischen Inhalt des Textes hin optimiert. Gemäß den Angaben der Forscher wurden Presstexte ausgeschlossen und die datenschutzrechtlichen Opt-Out-Hinweise in den txt-Dateien berücksichtigt und eingehalten. 60% der zu Trainingszwecken verwendeten Texte waren in Englisch und 40% in den 999 übrigen Sprachen. Die vielfältige Mehrsprachigkeit ist ein wichtiges Merkmal von Apertus für die mehrsprachige Schweiz im Vergleich zu (kommerziellen) Sprachmodellen aus den USA und Europa, die meist mit Englisch oder in der jeweiligen Landessprache konzipiert wurden.

Mit dem Aufkommen generativer KI-Modelle zeigen sich grundlegende Fragen nach der urheber- und datenschutzrechtlichen Zulässigkeit ihrer Trainingsmethoden

Fast gleichzeitig mit der Veröffentlichung von Apertus nahm im Schweizer Parlament eine rechtspolitische Debatte Fahrt auf, wie die urheberrechtlich geschützten Inhalte besser vor der kostenlosen Nutzung der Daten zu

KI-Trainingszwecken geschützt werden könnten. Die Regierung hat entschieden, im KI-Bereich einen sektorenbezogenen Ansatz der Rechtsanpassung zu verfolgen, anstatt ein KI-Gesetz zu erlassen.

Wie transparent und öffentlich zugänglich ist Apertus?

Gemäß den Angaben der Forscher in den Presseunterlagen ist Apertus ein durch und durch offenes und transparentes Datenmodell, da der Programmcode und die Gewichtung der Parameter in den neuronalen Netzen öffentlich zugänglich gemacht wurden. Nur Deepseek aus China ist auch eine bekannte offene KI-Anwendung.

Verglichen mit Apertus werden bei den sogenannten Open Weight Modellen wie Llama 3 von Meta oder GPT-OSS von OpenAI wesentliche Elemente nicht offengelegt:

1. Der Trainingscode – Die Skripts oder Frameworks, die zum Erstellen und Kuratieren des Trainingsdatensatzes verwendet werden.
2. Der Trainingsdatensatz – Der vollständige Datensatz, der für das Training verwendet wird, sofern dies rechtlich möglich ist. Eine Alternative, wenn die Verbreitung des Trainingsdatensatzes rechtlich nicht möglich ist.
3. Die umfassende Datentransparenz – Vollständige Details zur Zusammensetzung des Datensatzes, wie z.B. Quelldomänen, Bereinigungsverfahren oder Ausgleichstechniken.

Entspricht Apertus den rechtlichen Vorschriften des Datenschutzes und des Urheberrechts?

Ja, sagen die Forscher. Es wurden alle Vorgaben nicht nur des Schweizerischen Urheber- und Datenschutzrechts respektiert, sondern auch die Vorgaben der DSGVO und der Urheberrechtsschutzlinie der EU eingehalten. „Mit dieser Veröffentlichung möchten wir ein Musterbeispiel dafür geben, wie sich ein vertrauenswürdiges, souveränes und inklusives KI-Modell bauen lässt.“

Welche Anforderungen werden an das Training und an den Betrieb von Apertus gestellt?

Für das Training von Apertus wurde der Großrechner Alps benutzt. Dies schlug sich auch im Energieverbrauch nieder. Über drei Monate war die Hälfte der Kapazität des Großrechners Alps für das Training von Apertus reserviert. Mehr als 5 Gigawattstunden Strom wurden verbraucht. Das lässt sich mit dem Durchschnittsverbrauch von 1.500 Schweizer Haushalten während eines Jahres vergleichen. Legt man einen Elektrizitätspreis von 29 Rappen/Kwh zugrunde, so betragen die Energiekosten 1,5 Millionen Franken. Alps wird ausschließlich mit Wasserkraft betrieben, dadurch sind seine direkten CO₂-Emissionen minimal. Während in den USA ansässige Rechenzentren häufig Dieselgeneratoren für Notfälle besitzen, verfügt



The "Alps" research infrastructure. (Image: Marco Abram, CSCS) creativecommons.org/licenses/by-sa/3.0/deed.en

das CSCS über Notbatterien und eine Wasserreserve, was wesentlich umweltfreundlicher und nachhaltiger ist. Das große Sprachmodell erfordert für den Betrieb einen leistungsfähigen Computer mit 130 GB GPU-Speicher. Es kann zum Beispiel über die Swiss AI Platform der Swisscom (für Geschäftskunden) benutzt werden. Dort läuft Apertus auf dem Serversystem Nvidia DGX H100. Eine spezielle Hardware und mehrere Grafikkarten sind somit nötig. Die reduzierte Version begnügt sich hingegen mit 126 GB Arbeitsspeicher und soll auf einem guten Laptop oder auf einem handelsüblichen PC laufen.

Wer hat an Apertus mitgearbeitet – und wieviel hat es gekostet?

Zwischen 50 und 100 Forschende, Doktorierende und Ingenieure der ETH in Lausanne und Zürich waren an der Entwicklung von Apertus beteiligt. Der Supercomputer Alps kostete 100 Millionen Franken. Sein Betrieb ist öffentlich finanziert. Die Swiss AI Initiative läuft von 2025 bis 2028 mit einem Budget von total 20 Millionen Franken.

Welches sind die Leistungsmerkmale des Großrechners ALPS, auf dem Apertus trainiert wurde?

Die beiden Eidgenössischen Technischen Hochschulen der Schweiz (ETH/EPF) haben mit ALPS einen leistungsstärksten Supercomputer zur Verfügung, um große Sprachmodelle (LLMs) zu unterstützen. Alps bietet im Vollbetrieb zwanzig Mal so viel Rechenleistung wie sein Vorgänger Piz Daint und damit genug, um das Potential der KI voll auszuschöpfen.

In Sachen Leistungsfähigkeit liegt Alps weltweit auf Rang 7 – nur die USA, Finnland und Japan haben noch leistungsstärkere Rechner.

Was unterscheidet Apertus von ChatGPT und wie wird es eingesetzt?

Mit Apertus wollen die Forschenden der ETH ein offenes Fundament für Wissenschaft und Innovation schaffen, da ChatGPT ein Produkt mit Bezahlmodell ist. Apertus hingegen soll eine Grundlage sein, auf der Entwickler und Entwicklerinnen ihre eigenen Anwendungen aufbauen können. Die Forscher erwarten, dass Entwickler und Entwicklerinnen ihre spannendsten Ideen beisteuern. Zum ersten Mal eingesetzt wurde Apertus an den Hackathons der Swiss AI Weeks im September 2025. Vom 1. September bis 5. Oktober 2025 fanden im Rahmen der Swiss AI Weeks über 160 Veranstaltungen in 24 Schweizer Städten statt. Forscher, Firmen und Bürger und Bürgerinnen trafen dabei aufeinander. Ziel der Swiss AI Weeks ist, KI aus den Laboren herauszuholen und für breite Schichten anwendbar zu machen. Künstliche Intelligenz ist längst nicht mehr nur eine Sache für Experten und Expertinnen, sondern etwas, das alle betrifft.

Wie wird Apertus zugänglich gemacht und wie heruntergeladen?

Das Herunterladen und Einrichten des Modells ist für Fachpersonen und erfahrene Nutzerinnen und Nutzer relativ einfach. Die Hackathons der kommenden Swiss AI Weeks werden Entwicklern und Entwicklerinnen die erste Gelegenheit bieten, direkt mit Apertus zu experimentie-



creativecommons.org/licenses/by-sa/3.0/deed.en

ren, seine Fähigkeiten zu testen und Feedbacks zur Verbesserung zukünftiger Versionen zu geben.

Für Personen außerhalb der Schweiz wird zudem die Public AI Inference Utility Apertus als Teil einer weltweiten Bewegung für eine Öffentliche KI zugänglich machen. Nach dem Einrichten eines entsprechenden Nutzungskontos kann Apertus unter <https://apertus.rlp.de/slogin.php?ID=1> genutzt werden.

Die Parlamentsdebatte hat gezeigt, dass das Spannungsverhältnis groß ist: große Sprachmodelle (LLM) müssen mit großen Mengen an Daten trainiert werden, doch diese Daten sind nicht selten urheberrechtlich geschützt. Dabei steht den Urhebern eigentlich ein Ausschließlichkeitsrecht zu – zumindest, solange es keine spezifischen gesetzlichen Ausnahmen gibt.

Es ist daher von entscheidender Bedeutung, ein Gleichgewicht zwischen Innovation und Sicherheit, einschließlich der Rechte des Einzelnen, zu finden. Die Wissenschaft befürchtet, dass damit die Forschungsfreiheit eingeschränkt wird. Journalisten, Künstler, Medienunternehmen und Verwertungsgesellschaften verlangen einen konsequenteren Schutz urheberrechtlich geschützter Inhalte für das Training von Sprachmodellen und die Schließung allfälliger Lücken in der Gesetzgebung. Hier den Ausgleich zu finden, wird zu einer Herausforderung der Gesetzgeber in Europa werden.

Apertus ist als Modell ein «Impulsgeber für Innovationen und ein Mittel zum Ausbau der KI-Expertise in Forschung, Gesellschaft und Wirtschaft», betonten die Forscher auf der Pressekonferenz. Es bleibt zu hoffen, dass dies die Schweizer Bibliotheken als ein Fundament begreifen und

sich an den Hackathons der Swiss AI Weeks in den über 160 Veranstaltungen in 24 Schweizer Städten einbringen werden. Bis Anfang Oktober 2025 wurde Apertus 4.200.000-mal heruntergeladen. An Hackathons im Oktober in Schweizer Städten wurden bereits 50 Prototypen gezeigt, so zum Beispiel:

- Eine Lern-App gegen den Prüfungsstress. Man kann alte Prüfungen und Notizen hochladen und die App erstellt einen realistischen Lernplan nach den verbleibenden Stunden.
- Einen KI-Tutor für das Erlernen der tibetischen Sprache mit Textquellen.
- Ein KI-Tool zur Unterstützung von Flüchtlingsinterviews. Mitarbeitende können Transkripte hochladen, das System kategorisiert die Antworten automatisch und spart so bis zu 80 Prozent Zeit und Kosten.

Die weitere Zukunft von Apertus ist ungewiss. Die Projektbeteiligten hoffen, dass die Schweiz auf lange Sicht den Aufwand stemmen kann. Der Plan: jährlich ein verbessertes Modell und mit aktuellen Daten trainiert, eventuell in Kooperation mit Partnern in Europa. Das hängt aber auch von den finanziellen Mitteln der Bundesregierung in Bern und allenfalls von europäischen Forschungsförderprogrammen ab. **I**



Stephan Holländer

Basel
stephan@stephan-hollaender.ch