

Diese künstliche Intelligenz spricht auch Rätoromanisch

Die Schweiz zeigt, dass KI auch anders entwickelt werden kann als von den grossen IT-Konzernen – legal und transparent. Die Eidgenössischen Technischen Hochschulen in Zürich und Lausanne haben dafür zusammengespant. Hier ist ihr Rezept.

VON Ruth Fulterer (Text), Anja Lemcke (Illustrationen)

«Was ich nicht erschaffen kann, das verstehe ich nicht», soll der amerikanische Physiker Richard Feynman einmal gesagt haben. Und genau deshalb machten sich viele Computerwissenschaftler in den letzten Jahren Sorgen. Denn sprachliche künstliche Intelligenz (KI), also die Technologie hinter Chat-GPT, wird fast ausschliesslich von grossen Technologiefirmen gebaut. Forscher an Universitäten sind zunehmend abgehängt. Im Prinzip ist zwar bekannt, wie KI funktioniert. Doch sie zu bauen, ist sehr aufwendig. Und die Technologie-Firmen teilen immer weniger Details darüber, wie sie selbst vorgegangen sind.

In diesem Kontext ist das neue KI-Modell der Schweiz eine kleine Revolution. Es heisst «Apertus» und wurde am Dienstag um 10 Uhr morgens veröffentlicht. Und es ist zu hundert Prozent von Menschen gemacht, die über ihre Arbeit in der KI-Küche sprechen dürfen.

Wir haben diese einmalige Chance genutzt, um das Rezept für KI aufzuschreiben. Klar ist: Um einen Chatbot zu bauen, sind nicht nur enorme Ressourcen nötig, sondern auch Wissen, Teamgeist und starke Nerven.

1. Man nehme ethische Daten

Bettina Ursula Messmer trägt eine grosse Brille und spricht, obwohl sie nahe Zürich aufgewachsen ist, lieber auf Englisch über ihre Arbeit. Im Team ums Schweizer KI-Modell ist sie eine der wenigen Schweizerinnen. Sie hat drei Jahre in London gelebt. Nach dem Informatikstudium an der ETH Zürich arbeitete sie dort für die Datenanalyse-Firma Palantir. Dann kamen die Vorläufer von Chat-GPT. «Plötzlich passierte so viel im Bereich KI. Ich wollte verstehen, was da passiert. Ich wollte zurück in die Forschung.»

Seit 2022 ist Messmer Doktorandin an der ETH Lausanne (EPFL), in der Gruppe von Martin Jaggi. Er ist einer jener Forscher, die Ende 2023 den Plan verkündeten, ein Schweizer Sprachmodell zu bauen. Dass Messmer dabei sein würde, war von Anfang an klar. Denn sie hatte sich damit beschäftigt, wie man Sprach-Datensätze mit hoher Qualität herstellt. Und gute Datensätze sind die erste und vielleicht sogar die wichtigste Zutat für Sprach-KI.

Das Schweizer Modell sollte vielsprachig sein – und regelkonform. Im Kontext von Sprach-KI bedeutet das: Man hält sich ans Urheberrecht und nutzt keine Daten gegen den Willen der Hersteller. In anderen Branchen mag es üblich sein, solche Regeln zu beachten. Bei KI ist das anders. Denn in KI-Modelle fliessen Daten ein, die automatisiert aus dem Internet ausgelesen wurden. So war es auch bei Apertus. Doch vor der Verwendung wurden die Daten geprüft und gefiltert.

Messmer und ihre Kollegen. Sie suchten in Kleinarbeit riesige Datensätze zusammen und kontrollierten, woher die Daten

stammten. Sie filterten die Daten von Websites heraus, die einen Hinweis gegen die Verwendung für KI enthalten. «Das ist zwar aufwendig und auf seine Weise herausfordernd, richtet sich aber nach bestimmten Kriterien», sagt Messmer. Schwieriger sei die Aufgabe, problematische Aussagen aus Daten zu filtern wie Rassismus oder Antisemitismus.

Denn es ist schwer, per Computer normale Aussagen und Hassrede scharf zu trennen. Und wenn man zu viele Daten über eine Minderheit herausfiltert, könne die KI dann selbst in positiven Kontexten weniger gut Texte über die betreffende Gruppe erzeugen, erklärt Messmer. Auch wieder ein Problem. Ihr Team versuchte, einen Mittelweg zu finden. Ganz neutral sei eine KI nie. «Deshalb ist so wertvoll, dass wir offenlegen, wie wir vorgegangen sind. Das macht Verzerrungen sichtbar.»

Mehrere Monate dauerte die Arbeit, aus Internetdaten eine geeignete Quelle für die Schweizer KI zu machen. Immer wieder testete das Team in Experimenten, von welchen Daten das Modell wohl am besten lernen würde. Sie arbeiteten mit Annahmen. Denn der Trainingslauf eines grossen KI-Modells ist zu teuer, um ihn oft zu wiederholen und anzupassen.

Der Datensatz, aus dem Apertus entstanden ist, enthielt am Ende ungefähr 15 000 Milliarden Tokens, also Wörter oder Wortteile – etwa so viele wie in 19 Millionen Bibeln stecken. Mehr als 1000 verschiedene Sprachen kommen in den Daten vor – auch Rätoromanisch.

2. Man mische die Zutaten in einem Supercomputer

Maria Grazia Giuffreda ist Co-Direktorin des Schweizer Supercomputer-Zentrums CSCS in Lugano, sie erzählt viel und lebendig. Im Gespräch kommt bald der Stolz der Italienerin durch: «Ohne uns gäbe es kein Schweizer KI-Modell.»

Tatsächlich hat die Schweiz mit der «Alps» genannten Rechenmaschine im CSCS eine weltweit fast einzigartige Position: Sie besitzt seit 2024 einen grossen öffentlichen Supercomputer, der neueste Grafikprozessoren enthält und somit KI-tauglich ist. Wenn die wichtigste Zutat für KI Daten sind, ist ein riesiges Rechenzentrum die Küchenmaschine, sozusagen der Thermomix im KI-Rezept.

Für die meisten Länder wäre ein Projekt wie Apertus schon deshalb nicht möglich, weil ihnen dafür die Infrastruktur fehlt. Besonders für den ersten Teil der Entwicklung von Sprach-KI, das sogenannte Pre-Training, also Vortraining, sind viele vernetzte Prozessoren nötig, die gleichzeitig rechnen. Etwa ein Viertel der achtzig Ingenieure des CSCS waren monatelang vor allem damit beschäftigt, den neuen Supercomputer am Laufen zu halten und technische Probleme zu lösen.

Beim Pre-Training verarbeiten die Rechner die riesigen Datensätze und errechnen daraus die Parameter des KI-Modells. Das sind jene Variablen, die später verwendet werden, um die passende Antwort auf jede Frage zu errechnen. Apertus gibt es in zwei Versionen. Die grössere hat 70 Milliarden Parameter. Sie alle müssen durch Optimierung ausgerechnet werden. Man setzt dem Modell nach und nach die Trainingsdaten vor und lässt es

daraus lernen, vorherzusagen, welches Wort wahrscheinlich als Nächstes kommt.

Mehr als die Hälfte der Entwicklung von Alps war für das Training der KI reserviert. Der Supercomputer hat aber auch noch andere Nutzer: Meteo Schweiz etwa, Klimaforscher und Materialwissenschaftler. Auch sie brauchen Rechenpower. Das setzte der Grösse des Schweizer KI-Projekts Grenzen. Das grosse Apertus-Modell gehört zu den guten, offenen KI-Modellen – aber führende Modelle sind viel mächtiger. Die chinesische KI Deepseek ist zehnmal so gross wie Apertus.

Darüber, was der beste Moment beim Training von Apertus war, muss Giuffreda keine Sekunde nachdenken: «Wir wollten das Pre-Training des grossen Modells bis 23. Juli abgeschlossen haben. Ich verschickte täglich morgens um halb acht eine E-Mail mit dem Rechenfortschritt, um alle Mitarbeiter auf dem Laufenden zu halten.» Manchmal war es auch ein Rückschritt, und die Prognose hatte sich über Nacht verschlechtert.

«Und dann, am 25. Juli, schaute ich um Mitternacht in das Programm, und plötzlich stand da bei «Zahl der zu verarbeitenden Tokens»: null. Das war grossartig, das beste Gefühl!» Um 0 Uhr 22 hatte Giuffreda ihre Glückwünsche per E-Mail an das gesamte Team verschickt.

Der wichtigste grosse Schritt der Zubereitung war getan. Jetzt fehlten Verfeinerung und Würze.

3. Man bringe der KI bei, sich zu benehmen

Das Pre-Training eines KI-Modells ist nur der Anfang. Das erklärt der israelische ETH-Forscher Ido Hakimi mit einer Anekdote. Vor seiner Rückkehr an die Universität hat er unter anderem für Google gearbeitet. Die erste Sprach-KI, die er dort ausprobierte, war ein nur vortrainiertes Sprachmodell.

«Ich fragte: «Wie viel ist 2+2?» Rate mal, was die Antwort war.» Man muss nicht viel über KI und Pointen wissen, um zu vermuten: Wahrscheinlich war es nicht «4». Was Hakimi sagt, ist trotzdem überraschend. «Die Antwort war: «Wie viel ist 2+3?»» Der Grund: Im Pre-Training lernt die KI nur, wie ein Text wahrscheinlich weitergeht. Und in ihren Übungsdaten – etwa einem Schulbuch mit Rechenübungen – folgt auf eine Frage nicht immer eine Antwort, sondern oft auch die nächste Frage.

Ido Hakimi und sein ungarischer Kollege Barna Pásztor sollten aus einer Fortsetzungsmaschine, in der fast alles Wissen der Welt steckt, einen Gesprächspartner machen. Die beiden lachen viel und ergänzen ihre Sätze, als wir uns unterhalten. Künstliche Intelligenz lernt gutes Benehmen wie alles andere auch aus Beispielen. Fragende erwarten eine Antwort. Um das der KI beizubringen, haben Hakimi und Pásztor Datensätze mit Gesprächen gesucht.

Wenn die KI gelernt hat, einen Dialog zu führen, geht man zum nächsten Schritt über. Dafür braucht es Beispiele von guten und schlechten Antworten auf Fragen. Ein Algorithmus lernt daraus, die Qualität einer Antwort zu bewerten. Dann lässt man die KI viele Fragen und Antworten generieren. Der Prüfalgorithmus benotet die Antworten. Dadurch bekommt die KI Feedback und entwickelt so nach und nach Methoden, um gute Antworten zu geben.

So wird aus einer generischen KI-Masse ein fertiges Produkt mit ganz bestimmter Würze und Geschmack. Dazu gehört auch, der KI moralische Werte anzuerziehen. Sie muss wissen, wann sie besser nicht antworten sollte, zum Beispiel wenn jemand eine Bombe bauen will oder wenn Minderjährige pornografisches Material verlangen. Diese Grenzen gut umzusetzen, ist allerdings

schwierig. Immer wieder zeigen Skandale, dass auch Tech-Firmen damit kämpfen.

Die Forscher haben die Entwicklung des Modells immer wieder mit Tests geprüft, bei denen sie der KI etwa Programmieraufgaben stellten. Einmal verschlechterten sich die Fähigkeiten von Apertus auf unerklärliche Weise rapide. Nach langer Suche fanden sie den Grund: Ein einzelnes Leerzeichen im Code hatte dazu geführt, dass das KI-Modell nur noch Blödsinn von sich gab.

Für Hakimi ist das die Magie von KI: «In einem Moment sprichst du mit dem Modell, und es fühlt sich an, als wäre da eine Person. Und dann wird es durch ein einziges Leerzeichen komplett verwirrt. Umso erstaunlicher ist, dass die Sache funktioniert.»

Fast zärtlich reden die beiden über das KI-Modell, das sie mit erschaffen haben. An Apertus waren zwar insgesamt etwa hundert Personen beteiligt – doch das ist um einiges weniger als bei Tech-Firmen. Noch viel stärker kommt es da auf die Beiträge jedes Einzelnen an. Dieses Bewusstsein motiviert die Forscher.

4. Man lege fest, wo man eigentlich hinwill

Ich erreiche die Griechin Angelika Romanou kurz vor einer weiten Reise. Wenn dieser Artikel erscheint, wird sie schon in New York arbeiten, als Praktikantin beim Tech-Konzern Meta: Sie hilft mit, KI sicherer zu machen.

Sie sagt: «Was viele über meine Arbeit, und Forschung generell, nicht wissen, ist: Man ist kaum je damit beschäftigt, Lösungen für Probleme zu finden. Die meiste Zeit versucht man, das Problem zu definieren.» Die richtigen Fragen zu finden, war auch Romanous Aufgabe bei Apertus. Sie arbeitete im Evaluationsteam. Man könnte sagen, in der Testesser-Abteilung.

In der KI-Branche geht es immer um Benchmarks: Jemand stellt einen Test zusammen, zum Beispiel mit Programmierfragen, und alle Anbieter wetteifern darum, wer die besten Resultate erreicht. Es gibt Benchmarks zu Mathematik, Bilderkennung, juristischen Fragen. Für ein wichtiges Ziel, das sich das Schweizer KI-Projekt gesetzt hatte, fehlten aber gute Benchmarks: die Mehrsprachigkeit.

Bei Mehrsprachigkeit geht es nicht nur um korrekte Grammatik. KI lernt durch Sprache Weltwissen. Deshalb ist es relevant, aus welcher Sprache dieses Wissen kommt – und ob die KI das lokale Wissen auch in andere Sprachen übertragen kann. Romanou nennt ein Beispiel: «Stell dir vor, du bist in den Ferien in Griechenland, hattest zwei Drinks und möchtest von der KI wissen, ob du noch Autofahren darfst. Das Modell muss verstehen, dass es griechische Verkehrsgesetze auf Deutsch wiedergeben soll.»

Um diese Fähigkeit zu prüfen, sammelte sie mithilfe einer Community von KI-Forschern Fragen und Antworten in unterschiedlichsten Sprachen. Darunter waren viele afrikanische und asiatische Sprachen, die bei kommerziellen KI-Modellen oft schlecht funktionieren. Die Fragen und Antworten wurden wiederum in andere Sprachen übersetzt, um am Ende eine Sammlung an Fragen aus verschiedenen Ländern in allen möglichen Sprachen zu haben: Also einen Test, der echte Mehrsprachigkeit misst.

Wegen dieser Mehrsprachigkeit verzichteten die Forscher übrigens auf einen Hinweis auf die Swisness ihres Modells im Namen. Von Apertus soll nicht nur die Schweiz profitieren, sondern alle Forscher weltweit.

5. Man lasse die KI nachdenken, wie sie denken soll

Im Januar, als Apertus mitten im Training steckte, erlebte die Welt den Deepseek-Moment: Eine chinesische Firma hatte ein KI-Modell gratis angeboten, das gleich gut funktionierte wie die besten Modelle der kommerziellen Anbieter. Für die amerikanischen Tech-Firmen war das ein Schock. In Zürich freute sich einer: Eduard Durech.

«Es war schön, zu sehen, dass die Techniken, an denen ich arbeitete, auch im grossen Massstab funktionieren», sagt Durech. Er ist 27, in der Slowakei und in Kanada aufgewachsen und über die Verarbeitung von medizinischen Bildern zu KI gekommen. In Kanada hat er Computern beigebracht, anhand von Scans der Netzhaut Diabetes und Alzheimer zu erkennen. Die gleichen KI-Methoden wendet er heute auf Sprachdaten an. Es geht um sogenanntes Bestärkungslernen.

Beim Bestärkungslernen lernt der Computer durch Versuch und Irrtum. Indem er für gute Ergebnisse belohnt wird, bildet er nach und nach Strategien aus. Diese Technik auf Sprach-KI anzuwenden, hat Chatbots den letzten grossen Qualitätssprung beschert.

Durch diese Methode gebe Sprach-KI automatisch längere Antworten, zum Beispiel tausend statt hundert Worte, sagt Durech: «Wenn du ihm eine schwierige mathematische Frage stellst, beginnt es, verschiedene Methoden auszuprobieren, das Problem in Teilfragen einzuteilen, die Sache in Schritten anzugehen. Es simuliert Nachdenken.»

Die KI führt also eine Art reflektives Selbstgespräch, das ihre Antworten verbessert. Durech vermutete schon lange, dass bestärkendes Lernen hinter dem nächsten KI-Fortschritt stecken würde. Er knobelt seit letztem Herbst daran, die Methode für Sprach-KI zu adaptieren. Deepseek bestätigte ihm, dass sein Team auf dem richtigen Weg war.

Es dauerte aber eine Weile, bis Durech seine Methoden an Apertus einsetzen konnte. Denn seiner ist der letzte Schritt in der KI-Entwicklung. Erst seit Abschluss des restlichen Trainings kann sein Team auf grosse Computerressourcen zugreifen. Entsprechend dunkel sind die Ringe unter den Augen bei unserem Gespräch. Im Moment schlafe er nicht viel, sagt er.

Die Reasoning-Fähigkeiten kann man noch nicht ausprobieren – dafür hat die Zeit nicht gereicht. Aber noch im September soll das entsprechende Update veröffentlicht werden. Es ist die letzte Garnitur im KI-Rezept.

6. Man entlasse die KI in die Welt – und drücke die Daumen

Jetzt ist Apertus aufgetischt. Man kann das Modell als Chatbot ausprobieren oder es herunterladen und auf dem eigenen Computer laufen lassen. Auf den Internetplattformen Reddit und Hackernews diskutiert die KI-Community über das neue Modell. Die einen loben den offenen Ansatz, andere äussern sich enttäuscht, zum Beispiel über die Fähigkeiten der KI auf Finnisch und Bulgarisch.

Im technischen Bericht beschreiben die Forscher, wie sich Apertus in verschiedenen Benchmarks schlägt. Unter den transparenten KI steht Apertus demnach gut da, halboffene KI-Modelle von Meta und Google funktionieren aber besser. Doch ob Apertus als Erfolg oder als Flop in die Geschichte eingeht, hängt nicht nur von Rankings ab, sondern vom Interesse, mit dem die Wissenschaft das Modell annimmt – und von der Wirtschaft.

Die Entwickler hoffen, dass Apertus etwa für KMU eine Grundlage für eigene KI-Anwendungen sein kann – eine Chance, sich so von amerikanischen Anbietern unabhängig zu machen. Bei der Veranstaltungsreihe «Swiss AI weeks» soll die

interessierte Schweizer Bevölkerung dazu animiert werden, das Modell auszuprobieren.

Davon, wie Apertus ankommt, wird abhängen, wie viel die Schweiz weiter in offene KI investiert. Wird man ein noch grösseres Modell trainieren? Eines mit weiteren Fähigkeiten?

Die Forschenden sind jedenfalls motiviert, zu beweisen, dass es geht: Als vergleichsweise kleines Team von Professoren, Ingenieuren und Studenten künstliche Intelligenz zu entwickeln, die sich auf der Weltbühne nicht zu verstecken braucht.

«Man ist kaum je damit beschäftigt, Lösungen für Probleme zu finden. Die meiste Zeit versucht man, das Problem zu definieren.»

Angelika Romanou

Forscherin im Evaluationsteam von Apertus

Der Supercomputer «Alps» in Lugano erbrachte die nötige Rechenleistung, um die Schweizer KI namens Apertus zu entwickeln. Marco Abram