

Open Source AI

Zwischen Transparenz und Souveränität

Offene KI-Modelle gewinnen insbesondere im öffentlichen Sektor an Bedeutung. Für volle Transparenz braucht es aber auch Einsicht in die Trainingsdaten.



Damit Software als «Open Source» anerkannt wird, gibt es klare Kriterien: Es muss möglich sein, den Quellcode zu verwenden, zu studieren, zu verändern und zu teilen (vgl. «Open Source Software», S. 48). Die genauen Bedingungen werden durch Open-Source-Lizenzen definiert (vgl. «Beschaffung von Open Source Software», S. 20). Die Anwendung dieser Terminologie auf Künstliche Intelligenz (KI) ist noch nicht etabliert. Die gemeinnützige Organisation Open Source Initiative (OSI) definiert den Begriff «Open Source» für die Softwareentwicklung – vertritt aber die Auffassung, dass die gleichen Prinzipien auch auf KI angewendet werden sollten. Demnach müssen Datensätze, Code, Modellarchitektur und Modellparameter veröffentlicht werden, damit ein Modell als «Open-Source AI» gilt. Werden die Trainingsdaten oder der Trainingscode nicht veröffentlicht, bezeichnet die OSI diese Modelle als «Open Weights». Die meisten offenen Large Language Models (LLM), wie Llama, Gemma und GPT-OSS, gehören zur Kategorie der Open-Weights-Modelle.

Verschiedene Arten offener Modelle

Offene Modelle können dem öffentlichen Sektor einige Vorteile bieten, da sie zur Reduktion von Abhängigkeiten beitragen (vgl. «Digitale Souveränität», S. 46). Open-Weights-Modelle ermöglichen zwar den Betrieb auf eigener Infrastruktur, hinsichtlich künftiger Weiterentwicklungen sind die Nutzenden jedoch auf die Verfügbarkeit bereits trainierter Modelle angewiesen. Open-Source-Modelle erlauben dagegen das eigenständige Trainieren eines Modells. Von vollständiger Souveränität zu sprechen wäre jedoch auch in diesem Fall nicht zutreffend. Das Trainieren von State-of-the-Art-Modellen erfordert erhebliche Rechenleistung. Ist diese nicht vorhanden, besteht ein Abhängigkeitsverhältnis zu den Betreibern der Rechenzentren, auf der die Modelle trainiert werden.

Da sich Open-Weights- und Open-Source-Modelle auf eigener Infrastruktur betreiben lassen, können öffentliche Akteure unabhängig von (meist ausländischen) KI-Anbietern operieren. Durch das lokale Betreiben dieser Modelle lösen sich die meisten datenschutzrechtlichen Probleme, da die Daten die eigene Infrastruktur nicht mehr verlassen (vgl. «Rechtliche Aspekte von KI», S. 60). Open-Source-Modelle erfüllen zudem ein wichtiges gesellschaftliches Bedürfnis in Bezug auf Transparenz. Durch das Wissen darüber, mit welchen Daten und wie die Modelle trainiert wurden, lassen sich Vorurteile, Wertvorstellungen und Sicherheitsaspekte des Modells besser verstehen.

Schweizer Open-Source-LLM als Lösung?

2025 wurde im Rahmen der Swiss AI Initiative erstmals ein offenes, in der Schweiz trainiertes LLM (Apertus) veröffentlicht. Es entspricht den Kriterien von Open Source AI. Dies ist eine wichtige Grundlage für KI-Anwendungen im öffentlichen Sektor. Da sowohl die Expertise als auch die verwendeten Rechenzentren vollständig in der Schweiz angesiedelt sind und das Modell von Grund auf trainiert wurde, kann dieses Modell als souveräne Lösung betrachtet werden. Es wäre jedoch überraschend, wenn das Modell den State of the Art der proprietären Modelle konkurrenziert. Es ergibt sich ein gewisser Kompromiss: Vollständige digitale Souveränität lässt den Zugriff auf die besten Modelle nicht zu – und umgekehrt.

Unsere Empfehlungen



1. Apertus austesten

Das neue Sprachmodell bietet Potenzial für den öffentlichen Sektor. Es ist ratsam, dieses Modell in Pilotprojekten zu testen, um seine Leistungsfähigkeit, Anwendungsfälle und Limitationen besser zu verstehen.

2. Infrastrukturüberlegungen anstellen

Der Betrieb eigener KI-Modelle stellt spezifische Anforderungen an die Infrastruktur. Ein frühzeitiger Entwurf darüber, welche Modelle in welchem Umfang betrieben werden sollen, ermöglicht eine entsprechende Planung.

3. Anforderungen zum Datenschutz kennen

Es ist zu klären, ob bestimmte Informationen nur intern im eigenen Departement, nur im Inland oder auch im Ausland verarbeitet werden dürfen.

Mehr Informationen



Kontaktmöglichkeiten
und weitere Informationen
zu Open Source AI:
bfh.ch/ipst/public-sector-ai

Kontakt



Prof. Dr. Marcel Gygli

Professur KI im öffentlichen Sektor

marcel.gygli@bfh.ch

T +41 31 848 64 90



Leander Rankwiler

Wissenschaftlicher Mitarbeiter

leander.rankwiler@bfh.ch

T +41 31 848 32 04