



Optimisten sagen: SLMs sind in der Lage, die KI zu demokratisieren.

Wer sich mit künstlicher Intelligenz beschäftigt, kommt kaum um sogenannte Large Language Models (LLMs) herum. Vor allem die KI-Sprachmodelle Generative Pre-trained Transformer 3 (GPT-3) von OpenAI oder BERT von Google wurden in den letzten Jahren in den Medien rauf und runter gespielt. Immer öfter machen auch Small Language Models (SLM) auf sich aufmerksam. Wie der Name verrät, sind es kleine Sprachmodelle, und sie kommen mit deutlich weniger Parametern aus. Sie tragen dazu bei, KI-Technologien zugänglicher, flexibler und anwendungsorientierter zu machen und reduzieren die Abhängigkeit von großen Anbietern.

Bei der Recherche zu diesem Artikel zeigte sich jedoch auch die Tendenz, dass die Wirksamkeit oder der gezielte Einsatz von SLMs aktuell noch eine klar untergeordnete Rolle in der Wissenschaft spielen. Wie uns ein Insider wissen lässt, werden Forschende in der Computerlinguistik zunehmend danach bewertet, ob sie in ihren Studien die neuesten und leistungsfähigsten LLMs einsetzen. SLMs werden nicht beachtet, obwohl sie stärkere Berücksichtigung verdienen wür-

*SLMs werden
intelligenter,
datenschutzfreundliche
Anwendungen auf
Smartphones,
Wearables und IoT-
Geräten ermöglichen,
die eine
Echtzeitverarbeitung
ohne Abhängigkeit von
der Cloud erlauben.*

Udo Urbantschitsch

Senior Vice President bei Red Hat

den, gerade weil sie im Hinblick auf Aspekte wie Teilhabe und geringere Umweltbelastung deutliche Vorteile bieten.

ON-PREMISES EINSETZBAR. Small Language Models bieten eine Vielzahl von Anwendungsmöglichkeiten, insbesondere für kleine und mittelständische Unternehmen (KMUs), die von den energieeffizienten und ressourcenschonenden Eigenschaften dieser Modelle profitieren können. SLMs benötigen weniger Rechenleistung, was sie ideal für den Einsatz in Kundenservice-Chatbots oder automatisierten Textanalysen macht.

Beispielsweise können Unternehmen SLMs wie LLaMA 2 für interne FAQs oder IT-Support-Anfragen nutzen. Sebastian Schuster, Professor für Natural Language Processing an der Fakultät für Informatik der Universität Wien, hebt eine besondere Stärke kleiner Sprachmodelle hervor: „Sie können auf weniger rechenstarken Systemen verwendet werden, wie zum Beispiel auch auf mobilen Geräten.“ Man kann sie also lokal (on-premises) einsetzen, was den Datenschutz und die Sicherheit sensibler Daten erhöht. „SLMs könnten theoretisch besser