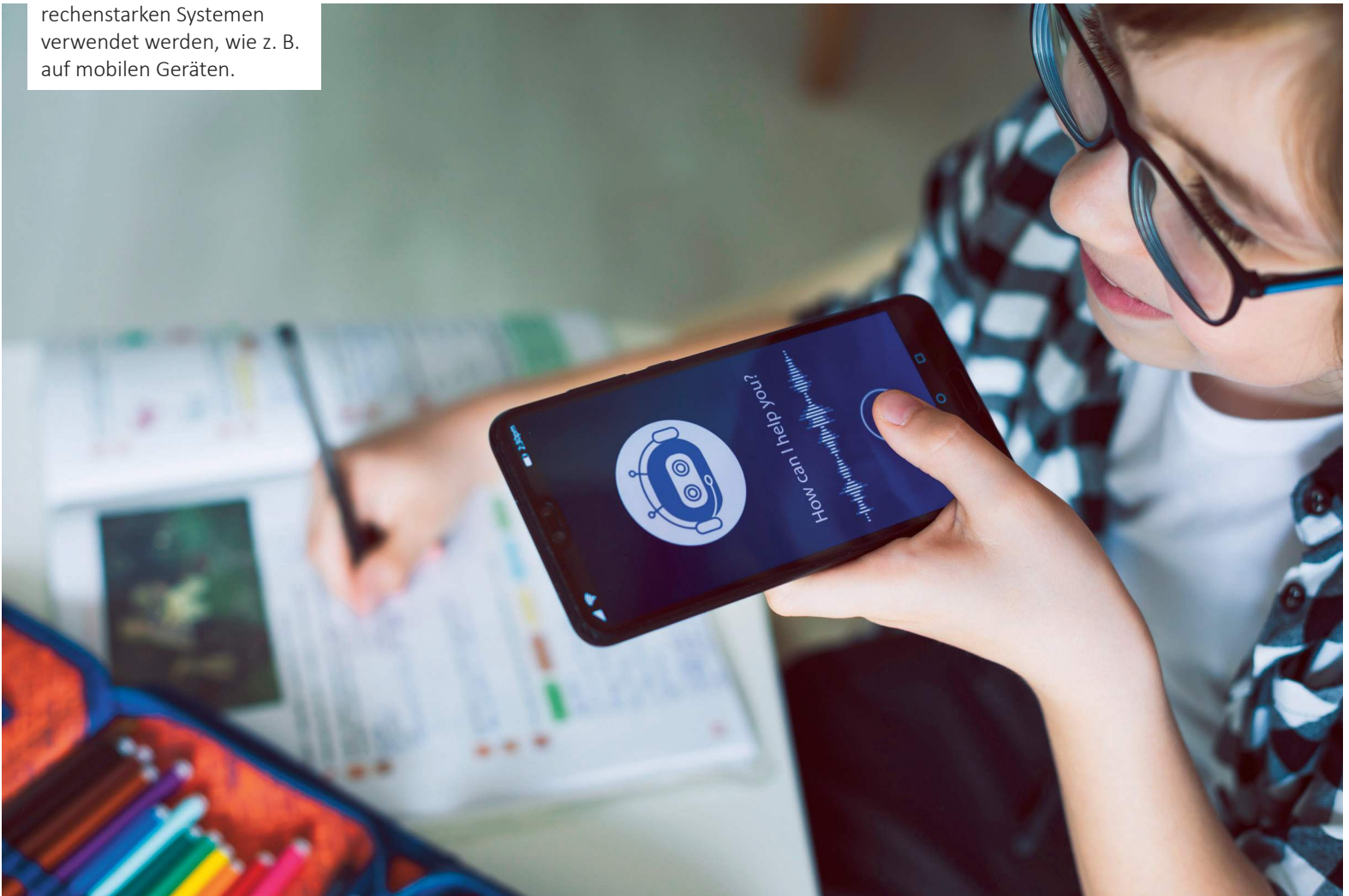


SLMs können auf weniger rechenstarken Systemen verwendet werden, wie z. B. auf mobilen Geräten.



auf individuelle Nutzer angepasst werden, da jeder Nutzer sein eigenes Modell verwenden könnte“, setzt Schuster fort. „Das würde zum Beispiel KI-Systeme ermöglichen, Texte im eigenen Stil zu verfassen oder Systeme, die Menschen mit Sprachbehinderungen beim Kommunizieren helfen könnten.“

An genau so einer Innovation arbeitet Moritz Grosse-Wentrup, Professor für Neuroinformatik an der Fakultät für Informatik. Er untersucht, inwiefern aktuelle Sprachassistenzsysteme, die auf vorgefertigten Piktogrammen basieren, durch fortschrittliche KI ergänzt werden können, um in neuen Situationen passgenaue Antworten vorzuschlagen. Zukünftig sollen Gehirn-KI-Schnittstellen die direkte, intuitive Kommunikation ohne bewusste Interaktion ermöglichen. „Wir entwickeln LLMs für Menschen mit Sprachbehinderungen und wollen den Sprachmodellen beibringen, für diese Menschen zu reden. Dafür benötigen die LLMs viele Informationen über den Menschen, den sie nachahmen“, so Grosse-Wentrup.

Hinsichtlich Datenschutz und Hackbarkeit ist das natürlich sensibel. „Deshalb ist es wichtig, dass diese Modelle, die wir entwickeln, lokal gehostet werden und unter Kontrolle der Nutzer stehen.“ Und dafür sind SLM ideal geeignet, weil sie keine großen Ressourcen brauchen, sondern auf kleinen Geräten lokal beim User laufen können, ohne dass die Daten in einer Cloud gehostet werden müssen. „In vielen Settings ist ein kleines Modell ausreichend, das über keinen großen Server laufen muss“, sagt Grosse-Wentrup. Ein Riesenvorteil hinsichtlich Datensicherheit und Privatsphäre.

ERSTE WAHL. SLMs können generell in datenschutzrelevanten Anwendungsfällen ihre Stärke ausspielen. „Sie eignen sich für Szenarien, die einen strengen Datenschutz und die Einhaltung gesetzlicher Vorschriften erfordern, wie zum Beispiel die Verarbeitung sensibler Patienten- oder Finanzdaten“, führt Udo Urbantschitsch aus, Senior Vice President des Open-Source-Unternehmens

Red Hat. Der internationale Top-Experte sieht zahlreiche Felder, in denen SLMs die optimale Wahl sein können, darunter auch domänenspezifische Anwendungen. „Sie eignen sich hervorragend für Aufgaben, die tiefes Fachwissen oder kontextbezogenes Verständnis in Bereichen wie Gesundheitswesen, Recht, Finanzen und Kundensupport erfordern, wo Genauigkeit und Relevanz entscheidend sind.“

Aber auch bei Echtzeit- und Niedriglatenz-Anforderungen. „SLMs bieten eine schnellere Inferenz und Reaktionsfähigkeit, was sie perfekt für Echtzeitanwendungen macht.“ Nicht zu verachten ist die Kosteneffizienz. „Für Unternehmen, die ihre Infrastruktur- und Betriebskosten minimieren möchten, bieten SLMs eine praktische Alternative zu großen, cloudabhängigen Modellen.“

VON INDUSTRIE BIS BILDUNG. Auch in der Industrie können SLMs eingesetzt werden und etwa Maschinenprotokolle analysieren und frühzeitig Anomalien erkennen,