MedWiss.Online

Gefährliche Chatbots: Prof. Stephen Gilbert fordert Zulassung als Medizinprodukt

LLM-basierte generative Chat-Tools wie ChatGPT oder MedPaLM von Google haben großes medizinisches Potenzial, ihre unregulierte Verwendung im Gesundheitswesen birgt jedoch inhärente Risiken. Der neue Nature-Medicine-Beitrag von Prof. Stephen Gilbert et. al. befasst sich mit einem der drängendsten internationalen Probleme unserer Zeit: Wie lassen sich Large Language Models (LLMs) im Allgemeinen und im Gesundheitsbereich im Besonderen regulieren?

Große Sprachmodelle sind neuronale Netzwerke mit bemerkenswerten Konversationsfähigkeiten. Sie erzeugen menschenähnliche Reaktionen und führen interaktive Gespräche. Allerdings generieren sie oft äußerst überzeugende Aussagen, die nachweislich falsch sind oder unangemessene Antworten liefern. Heutzutage ist es nicht möglich die Qualität, Stichhaltigkeit oder Zuverlässigkeit der gegebenen Antworten zu überprüfen. „Diese Chatbots sind unsichere Werkzeuge, wenn es um medizinische Beratung geht und es ist notwendig, neue Rahmenbedingungen zu entwickeln, die die Patientensicherheit gewährleisten“, sagte Prof. Stephen Gilbert, Professor für Medical Device Regulatory Science am Else Kröner Fresenius Zentrum für Digitale Gesundheit an der TU Dresden.

Herausforderungen bei der behördlichen Zulassung großer Sprachmodelle

Die meisten Menschen informieren sich online über ihre Symptome, bevor sie ärztlichen Rat einholen. Suchmaschinen spielen im Entscheidungsprozess eine wichtige Rolle. Die bevorstehende Integration von LLM-Chatbots in Suchmaschinen könnte das Vertrauen der Benutzer:innen in die Antworten eines Chatbots, der menschliche Konversationen nachahmt, erhöhen. Es hat sich jedoch gezeigt, dass LLMs äußerst gefährliche Informationen liefern können, wenn sie mit medizinischen Fragen konfrontiert werden. Der zugrundeliegende Ansatz von LLM enthält kein Modell einer medizinischen „Grundwahrheit“, was gefährlich ist. Ground Truth ist eine Prozessmethode, die sicherstellt, dass die der Analyse zu Grunde gelegten Daten aktuell, präzise, vollständig und bestens gepflegt sind. LLMs mit Chat-Schnittstelle haben schon schädigende medizinische Antworten erzeugt und wurden bereits in Versuchen an Patienten, ohne deren Einverständnis, eingesetzt. Nahezu jeder medizinische LLM-Anwendungsfall erfordert in der EU und den USA eine behördliche Kontrolle. LLMs mit nachvollziehbaren Ergebnissen, geringer Verzerrung, vorhersagbar, korrekt und mit überprüfbaren Ergebnissen gibt es derzeit nicht. In diesem Nature-Medicine-Artikel beschreiben die Autoren die begrenzten Szenarien, in denen LLMs unter den aktuellen Rahmenbedingungen Anwendung finden könnten. Sie beschreiben, wie LLM-basierte Tools entwickelt werden können, die als medizinische Geräte zugelassen werden könnten, und sie untersuchen die Entwicklung neuer Rahmenbedingungen für die Sicherheit der Patienten. „Aktuelle LLM-Chatbots erfüllen nicht die wichtigsten Prinzipien für KI im Gesundheitswesen, wie Voreingenommenheitskontrolle, Erklärbarkeit, Aufsichtssysteme, Validierung und Transparenz. Um sich ihren Platz im medizinischen Repertoire zu verdienen, müssen Chatbots für eine höhere Genauigkeit konzipiert werden, wobei Sicherheit und klinische Wirksamkeit nachgewiesen und von den Aufsichtsbehörden genehmigt werden müssen“, schließt Prof. Gilbert.