

Steffen Brandt
opencampus.sh
Fraunhoferstr. 13
Wissenschaftszentrum Kiel
24118 Kiel

Schleswig-Holsteinischer Landtag Umdruck 20/2904

An
Herrn Claus Christian Claussen
Vorsitzender des Wirtschafts- und Digitalisierungsausschusses
Schleswig-Holsteinischer Landtag
Düsternbrooker Weg 70
24105 Kiel

**Betreff: Schriftliche Anhörung des Wirtschafts- und
Digitalisierungsausschusses**

Thema: Entwurf eines Gesetzes zur Änderung des Landeskrankenhausgesetzes
Drucksache 20/1764

Sehr geehrter Herr Claussen,

als Fachexperte im Bereich Künstliche Intelligenz (KI) durch meine Tätigkeit bei opencampus.sh und als Mitglied des „Virtuellen Kompetenzzentrum: Künstliche Intelligenz und wissenschaftliches Arbeiten“ (VK:KIWA) begrüße ich die Gelegenheit, zur geplanten Änderung des Landeskrankenhausgesetzes in Schleswig-Holstein Stellung zu nehmen. Mein Fokus liegt auf der Bedeutung des Zugangs zu hochwertigen und umfangreichen Datensätzen für die Entwicklung und das Training leistungsfähiger KI-Modelle.

Allgemeine Bewertung des Gesetzesentwurfs:

Die Neufassung des § 38 des Landeskrankenhausgesetzes, die eine verbesserte Nutzung von Gesundheitsdaten für Forschungszwecke vorsieht, ist ein wichtiger Schritt, um die Entwicklung besserer KI-Systeme im Gesundheitssektor zu ermöglichen und Schleswig-Holstein ein Teil dieser Entwicklung werden zu lassen.

Eine verantwortungsvolle und datenschutzkonforme Nutzung von Daten aus Bereichen wie dem Gesundheitswesen sind hierbei Grundvoraussetzung.

Spezifische Anmerkungen:

Der Zugang zu großen, qualitativ hochwertigen und vielfältigen Datenmengen ist eine entscheidende Voraussetzung für den Erfolg von KI-Systemen in nahezu allen Anwendungsbereichen. In der Vergangenheit hat sich gezeigt, wie stark der Fortschritt der KI-Entwicklung mit der Verfügbarkeit entsprechender Datensätze zusammenhängt:

(1) **Bildererkennung:** Die Veröffentlichung des ImageNet-Datensatzes mit über 14 Millionen annotierten Bildern aus über 20.000 Kategorien im Jahr 2009 markierte einen Meilenstein. Neuronale Netze, die auf diesem Datensatz trainiert wurden, erzielten einen Durchbruch in der Objekterkennung und machten den Weg frei für viele praktische Anwendungen der Computervision.

(2) **Sprachverarbeitung:** "The Pile" ist ein riesiger Datensatz mit über 825 Gigabyte an Rohdaten aus unterschiedlichsten Online-Quellen, der 2020 von EleutherAI veröffentlicht wurde. Auf Basis dieses Datensatzes konnten erstmals leistungsfähige Open-Source-Sprachmodelle wie GPT-Neo entwickelt werden und der Datensatz gab damit den Startschuss für eine neue Open Source-Bewegung im Bereich der Sprachmodelle mit aktuell wöchentlichen Neuerungen und Verbesserungen.

(3) **Bildgenerierung:** Der 2021 von LAION veröffentlichte LAION-400M Datensatz mit 400 Millionen englischen Bild-Text-Paaren bildete die Grundlage für die Entwicklung von Stable Diffusion und anderen leistungsstarken Text-zu-Bild-Modellen im Bereich der generativen KI.

Die Beispiele verdeutlichen, dass der Zugang zu großen, vielfältigen und aussagekräftigen Datensätzen ein kritischer Erfolgsfaktor für Innovationen im KI-Bereich ist.

Mit freundlichen Grüßen
Steffen Brandt