

OpenAI gewährt API-Zugriff auf GPT-3

Die gigantische Sprach-KI GPT-3 ist ab sofort über ein öffentliches API verfügbar. Das neuronale Netz schreibt angefangene Texte stillsicher weiter. Entwickler können dies nun in eigene Anwendungen einbauen, ohne sich aufwendig bei OpenAI bewerben zu müssen.

Das KI-Unternehmen OpenAI hat bereits 2020 ein unglaublich großes neuronales Netz namens „Generative Pre-trained Transformer 3“ (GPT-3) vorgestellt. Das Netz nutzt 175 Milliarden Parameter und hat beim Training terabyteweise Texte aus dem gesamten Internet gelesen. GPT-3 verblüfft seitdem Experten mit stilsicheren und fehlerfreien Texten in mehreren Sprachen, einschließlich Programmiersprachen. Jetzt ist der Zugriff auf GPT-3 für die Allgemeinheit freigegeben.

Transformer sind neuronale Netze, die Sequenzen als Eingabe lesen und neue Sequenzen ausgeben können. Beim Berechnen der Ausgabe können sie ihre Aufmerksamkeit auf bestimmte Teile der Eingabe richten und dabei lernen, eine interne Vektor-Repräsentation der Bedeutung der Eingabe gezielt an den Kontext anzupassen. Sie ähneln den älteren Long-Short-Term-Memory-Netzen (LSTM), lassen sich aber parallel berechnen und besser skalieren.

Der nächstliegende Anwendungsfall einer solchen KI ist Sprache: Sätze sind Sequenzen aus Wörtern, die die Transformer lesen. Die Ausgabe ist Wort für Wort ein neuer Satz. Ein solches Sprachmodell wird darauf trainiert, einen angefangenen Text weiterzuschreiben, indem es stets das beste nächste Wort berechnet.

Externen Programmierern gewährte OpenAI bislang nur selten Zugriff auf das Sprachmodell. Die Zurückhaltung begründete das Unternehmen damit, dass ein so gutes Sprachmodell leicht für Fehlinfor-

mationskampagnen, Spam oder Betrug missbraucht werden kann.

OpenAI hat nun intern genügend Sicherheitsmaßnahmen eingebaut, dass das Risiko für Missbrauch der KI nach Meinung des Unternehmens akzeptabel gering ist. Die Registrierung für das API ist nun öffentlich und ohne Warteliste freigegeben. Programmierer, die GPT-3 in ihren Anwendungen nutzen möchten, müssen Richtlinien zustimmen, die zum Beispiel Betrug und politische Einflussnahme untersagen sowie die erlaubten Anwendungsgebiete beschränken. Für viele Diktaturen ist der Zugriff ganz gesperrt. Ein Content-Filter stuft Texte als „safe“, „sensitive“ oder „unsafe“ ein. Bezahlt wird der API-Zugriff pro 1000 Wörter oder Satzzeichen (Tokens) mit Preisen zwischen 0,08 und 6 US-Cent, je nachdem welche Variante der KI man nutzt.

Ein Sprachmodell gegen unethische Verwendung abzusichern ist so schwierig, dass trotzdem mit Missbrauch zu rechnen ist. Anfang dieses Jahres haben Forscher bereits nachgewiesen, dass GPT-3 aus seinen Trainingsdaten Vorurteile über Muslime gelernt hat, die es auch in eigenen Texten wiedergibt. Das Replizieren von Vorurteilen aus dem Datensatz ist ein Problem, mit dem fast alle Sprach-KIs zu kämpfen haben. Da GPT-3 aber eloquenter schreibt als viele kleine neuronale Netze, sind seine Texte nur sehr schwer als das Werk einer Maschine erkennbar. Eine Quellenprüfung für Texte im Internet wird also immer wichtiger, auch weil GPT-3 nicht alleine bleiben wird. Das Start-up AlephAlpha arbeitet bereits an einem ähnlich großen Sprachmodell, genauso wie Microsoft und Nvidia, die ihrem Netz den einschüchternden Namen „Megatron-Turing Natural Language Generation Model“ (MT-NLG) gegeben haben – Megatron ist ein Bösewicht aus dem Transformers-Franchise. (pmk@ct.de)

Per-model prices			
Ada Fastest \$0.0008 <small>/ 1K tokens</small>	Babbage \$0.0012 <small>/ 1K tokens</small>	Curie \$0.0060 <small>/ 1K tokens</small>	Davinci Most powerful \$0.0600 <small>/ 1K tokens</small>
Multiple models, each with different capabilities and price points. Ada is the fastest model, while Davinci is the most powerful.			
<small>Prices are per 1,000 tokens. You can think of tokens as pieces of words, where 1,000 tokens is about 750 words. This paragraph is 35 tokens.</small>			

Für den Zugriff auf GPT-3 zahlt man pro Wort. Die hohen Preise ergeben sich aus der immensen Rechenleistung, die GPT-3 verschlingt.