



KI BUNDESVERBAND

Large European AI Models (LEAM)

Positionspapier zum Aufbau großer europäischer KI-Modelle

Mai 2021





1. Zusammenfassung

Aktuelle Fortschritte bei der Entwicklung großer KI Modelle haben die Chance, den gesamten Markt für Künstliche Intelligenzen zu verändern. Diese Fortschritte wurden bislang vor allem von amerikanischen Unternehmen vorangetrieben. Wir haben die Befürchtung, dass europäische Akteure hinter die amerikanischen zurückfallen und mit den schnellen Innovationszyklen nicht mehr mithalten können. Mittelfristig erwarten wir, dass das europäische KI-Ökosystem von amerikanischen Modellen abhängig wird. Wir haben diese Entwicklung bereits in anderen Digitalsektoren wie Suchmaschinen, mobilen Anwendungen oder Cloud Anwendungen erlebt.

Wie bei diesen Beispielen werden amerikanische Lösungen auch bei der Entwicklung von KI keine Rücksicht auf europäische Werte legen. Wir befürchten im Besonderen Defizite im Bereich Datenschutz. Vor allem kleinere Sprachfamilien werden unter dieser Entwicklung leiden, da nicht zu erwarten ist, dass amerikanische Unternehmen Produkte für diese entwickeln. Um die digitale Souveränität der EU zu gewährleisten, braucht die Europäische Union daher eine mutige Initiative zur Entwicklung großer KI Modelle.

2. Ausgangssituation

Seit circa 1,5 Jahren gewinnt *Language Modelling* (LM) im Kontext des *Natural Language Processing* (NLP) zunehmend an Bedeutung. LM umfasst die Verwendung verschiedener statistischer und probabilistischer Techniken zur Bestimmung der Wahrscheinlichkeit, dass eine bestimmte Wortfolge in einem Satz vorkommt. Sprachmodelle analysieren Textdaten, um eine Grundlage für ihre Wortvorhersagen zu schaffen. Sie werden in NLP-Anwendungen verwendet, insbesondere in solchen, die Text als Ausgabe erzeugen. Einige dieser Anwendungen umfassen die maschinelle Übersetzung und die Beantwortung von Sachfragen.

Ein aktuelles Projekt des US-amerikanischen KI-Unternehmens *OpenAI* wird als möglicher revolutionärer Durchbruch im Bereich LM gewertet. *OpenAI* hat für die



Erstellung des *Generative Pre-trained Transformer 3* (GPT-3)¹ ein neuronales Netz mit gewaltigen Datenmengen trainiert. Es wurde ein erheblicher Teil der im Internet frei verfügbaren Texte sowie Wikipedia und einige weitere Quellen genutzt, um ein Modell zu erstellen, das vielfältig eingesetzt werden kann: u.a. für semantische Suche, Übersetzungen, Textgenerierung, die Zusammenfassung von Text, etc.

Eine Besonderheit dieser Modelle ist ihre Fähigkeit zur Generalisierung von Text. Ohne spezielle Trainingsdaten können diese KI-Systeme komplexe "intelligente" Aufgaben flexibel lösen. Damit werden neue Möglichkeiten geschaffen, die keine bisherigen Modelle anbieten können. Die deutsche Presse hat die Entwicklung mit großer Aufmerksamkeit verfolgt.^{2,3,4}

Neben der Nutzung eines riesigen Korpus an Trainingsdaten erforderte die Erstellung von GPT-3 enorme Rechenkapazitäten (nach Angaben der Entwickler ca. 1000 PFLOPS Stunden was etwa 355 GPU-Jahren entspricht). Dies würde externen Berechnungen zufolge im Rahmen von Cloud Angeboten zwischen 5 und 20 Millionen USD kosten⁵ – lediglich für einen Durchlauf des Model-Trainings. Die Hardwareanforderungen sind dabei dergestalt, dass Standard GPU Rechenzentren nicht geeignet sind. Diese speziellen technischen Anforderungen müssen bei der Planung der Rechenzentren bereits berücksichtigt werden.

3. Ausblick und Relevanz

OpenAI hat mit GPT-3 bewiesen, dass man auf Basis großer Datenmengen und enormer Rechenkapazitäten in der Lage ist, führende, hochperformante KI-Modelle zu entwickeln. Gleichzeitig verdeutlicht der Wechsel von *OpenAI* von einer non-profit zu einer for-profit Organisation sowie die Umstellung des Geschäftsmodells auf ein kommerzielle Verwertung von GPT-3 das sich wandelnde Marktumfeld und die daraus resultierenden Abhängigkeiten.

¹<https://arxiv.org/abs/2005.14165>

²<https://www.spiegel.de/netzwelt/web/gpt-3-die-eloquenteste-kuenstliche-intelligenz-der-welt-a-dd3b3423-d214-4a2f-bc51-d51a2ae22074>

³<https://mixed.de/openai-gpt-3-das-ist-die-derzeit-maechtigste-sprach-ki/>

⁴<https://www.welt.de/wirtschaft/article223415594/GPT-3-Europa-muss-dieses-KI-Projekt-kopieren-sonst-verliert-es-den-Anschluss.html>

⁵<https://venturebeat.com/2020/06/11/openai-launches-an-api-to-commercialize-its-research/>



GPT-3 ist dabei nur ein erster Schritt in diese Richtung. Weitere zum Teil noch umfangreichere KI-Modelle werden in verschiedensten Bereichen folgen, u.a. etwa auch in der Bildverarbeitung.

Beispiele hierfür sind die im Januar 2021 von *OpenAI* vorgestellten Modelle DALL-E⁶ und CLIP⁷.

Wegen des erforderlichen Aufwands, vor allem an Rechenkapazität, können diese Modelle nur von finanz- und ressourcenstarken Unternehmungen umgesetzt werden. Dies wird mittelfristig dazu führen, dass zentrale KI-Funktionalitäten von einigen wenigen Marktteilnehmern bereitgestellt und diese von Anwendern über den Aufruf von Programmierschnittstellen in der Cloud genutzt werden. Das innovative deutsche KI-Ökosystem wird in der Konsequenz vom Zugang zu diesen Modellen abhängig.

GPT-3 ist nur per API verwendbar. Deutsche Firmen müssten also z.B. eigene Daten und Kundendaten derzeit über den Atlantik schicken, um das Modell verwenden zu können. Das ist für viele Firmen aufgrund des Datenschutzes nicht realisierbar.

4. Herausforderung

Es werden mittelfristige Monopole oder Oligopole für spezifische Bereiche der KI-Anwendungen entstehen. Diese werden sich zusätzlich dadurch verstärken, dass die Anbieter, die von den Kunden über die Schnittstellen bereitgestellten Daten zur Erweiterung des Trainingskorpus nutzen werden. Dies wird zu einer kontinuierlichen Verbesserung des Gesamtsystems beitragen ähnlich wie jede Suchanfrage den Vorsprung von *Google* vergrößert. Eine derartige Nutzung ist meist in den Geschäftsbedingungen der Anbieter vereinbart.

In der Konsequenz werden die Angebote der großen Anbieter, aufgrund ihrer qualitativen Überlegenheit auch wirtschaftlich uneinholbar werden. Sie werden ähnlich wie *Google* im Bereich der Internet-Suchmaschinen eine Marktmacht aufbauen, die zu einer einseitigen Abhängigkeit von Unternehmen und der gesamten Wirtschaft führt.

⁶<https://openai.com/blog/dall-e/>

⁷<https://openai.com/blog/clip/>



5. Handlungsbedarf

Um diese Situation auch unter der Zielsetzung einer europäischen digitalen Souveränität zu vermeiden, ist es erforderlich, zeitnah mit der Konzeption, Entwicklung und dem Aufbau der dafür erforderlichen Infrastruktur sowie den dafür notwendigen Strukturen zu beginnen.

Ziel ist der Aufbau eines deutschen bzw. europäischen KI-Ökosystems zur Entwicklung, Bereitstellung, Betrieb, Integration und Validierung besonders leistungsfähiger KI-Modelle. Dabei müssen europäische Werte und Standards in den Bereichen Datenschutz, Diskriminierungsfreiheit oder Offenheit genauso Berücksichtigung finden wie die Nachhaltigkeit der Entwicklung und des Betriebs.

Die erforderlichen Aktivitäten lassen sich in folgende zentrale Bereiche aufteilen:

1. Sammlung und Erstellung von umfassenden Trainingsdatensätzen (sowohl konventioneller Natur als auch unter Einbeziehung synthetisch⁸ erzeugter Daten), insbesondere auch für "typisch" europäische Domänen (siehe KI Strategie der EU) wie Medizin, Maschinen- und Fahrzeugbau, Bildung und generelle B2B-Szenarien. Die Trainingsdatensätze werden aus unterschiedlichen Quellen bereitgestellt, wie z.B. Crawling des (deutschsprachigen) Internets und der Bereitstellung von Daten aus Unternehmen und Institutionen über geeignete Plattform - unter Berücksichtigung aller rechtlichen Rahmenbedingungen und deren Umsetzung;
2. Fokussierte Forschung und Entwicklung im Bereich großer KI-Modelle;
3. Bereitstellung der erforderlichen Rechenkapazitäten in leistungsfähigen und nachhaltigen digitalen Infrastrukturen;
4. Vorbereitung von Integration und Partizipation des deutschen/EU Innovationsökosystems;

⁸<https://amp.welt.de/wirtschaft/article221217488/KI-Wenn-Daten-das-neue-Oel-sind-dann-ist-China-das-neue-Saudi-Arabien.html>



5. Entwicklung von Methoden, Benchmark-Datensätzen und Kriterien zur Validierung europäischer Werte;
6. Entwicklung eines Konzepts zur organisationalen Ausgestaltung der Entität, die diese Aktivitäten koordiniert.

INNOVATIONSPLATTFORM

Auf einer Innovationsplattform sollen die genannten Aktivitäten koordiniert und implementiert werden. Die Erstellung von Trainingsdatensätzen erfolgt im Bereich der Sprache und weiteren Einsatzmöglichkeiten großer KI-Modelle auf nationaler Ebene. Eine zentrale Stelle koordiniert bereits existierende Forschungsprojekte und stößt neue Initiativen zur Datensammlung an.

Die grundlegenden Algorithmen sind zum Teil bereits verfügbar, auch als *open source*. Weitere Forschung wird bereits in Deutschland an verschiedensten Einrichtungen betrieben. Die relevanten Aktivitäten und Ergebnisse werden koordiniert und konsolidiert sowie durch weitere Forschungsprojekte ergänzt.

Die Bereitstellung von erforderlichen Rechenkapazitäten könnte beispielsweise über das geplante IPCEI (Important Project of Common European Interest) zum Thema Industrial Cloud erfolgen. LEAM kann zudem als Leuchtturmprojekt für GAIA-X fungieren und entsprechende Synergien fördern.

Die großen KI-Modelle müssen bezüglich Datenbereitstellung (z.B. Anti-Bias-Optimierung) und ihrer Algorithmen stetig weiterentwickelt werden. Dies setzt die Bereitstellung und kontinuierliche Weiterentwicklung entsprechender Infrastruktur voraus. Der bereits hohe und stark wachsende Energiebedarf für eine solche Infrastruktur verlangt, dass hohe Anforderungen in Bereich Energieeffizienz und Nachhaltigkeit erfüllt werden.

6. Lösungsansatz

In einem ersten Projekt ist ein Detailkonzept zu erstellen, das u.a. die folgenden Elemente enthält:



1. Projektpartner;
2. Aufbau der Infrastruktur;
3. Datensammlung;
4. algorithmische Entwicklungsarbeiten;
5. Organisation und Prozesse;
6. Definition von Teilprojekten;
7. Finanzierungen.

Beteiligte eines entsprechenden Konzeption Projekts sollten u.a folgende Entitäten sein:

1. KMUs und KI-Unternehmen;
2. Industriepartner und KI-Anwender;
3. Vertreter aus dem Umfeld der GAIA-X Initiative;
4. Cloud-Computing Partner (z.B. *Cloud&Heat* und *Ionos*) zur Bereitstellung der Rechenkapazitäten;
5. DE-CIX (leistungsfähige Verbindung verschiedener Cloud-Standorte);
6. Universitäre Forschungseinrichtungen.

Auf Basis des Detailkonzepts wird eine modularisierte phasenweise Planung zur Realisierung des Vorhabens durchgeführt.

7. Fazit

GPT-3 hat aufgezeigt, dass große KI-Modelle enorm leistungsfähig sind und die Anwendung von KI in vielen Bereichen dominieren werden.

Aufgrund des erforderlichen Ressourcen- und Kapitaleinsatzes können diese Modelle derzeit nur zentral, von wenigen entsprechend ausgestatteten Unternehmen bereitgestellt werden. Es besteht die Gefahr, dass sich in diesem Bereich Monopole und Oligopole herausbilden, die zu einer einseitigen technologischen und wirtschaftlichen Abhängigkeit führen und dass ein wesentlicher Teil der Wertschöpfung in anderen Regionen realisiert werden wird.



Zum Erhalt der digitalen Souveränität in Europa ist der umgehende Start von Maßnahmen für den Aufbau eines offenen Service für große KI-Modelle (Arbeitstitel: *LEAM*) erforderlich.

Jörg Bienert
KI Bundesverband e.V.

Prof. Alexander Löse
Beuth Hochschule

Ronny Reinhardt
Cloudandheat GmbH

Johannes Bubenzer
2txt GmbH

Valeria Klassen
Mediengruppe RTL

Prof. Kristian Kersting
TU Darmstadt

Claudia Pohlink
Telekom T-Labs

Corina Apachite
Continental

Prof. Dominik Michels
Kaust University

Hauke Timmermann
Eco Verband e.V.