

Hochschule Stralsund

Exposé

**Für das Modul „Scientific Circle for Bachelor Thesis“
im Studiengang Wirtschaftsinformatik**

Name: Jonas Bevernis

Matrikelnummer: 16859

Erstgutachter: Prof. Dr. Thomas Wengerek

Stralsund, den 05.06.2020

Inhaltsverzeichnis

Inhaltsverzeichnis	2
1 Einleitung	3
2 Grundlagen	4
2.1 Neuronale Netze/Deep Learning	4
2.2 Texterzeugungssysteme	4
2.3 GPT	4
2.4 OpenAI	4
3 Forschungsziel	5
4 Forschungsmethodik	5
5 Titelvorschlag	6
6 Literaturrecherche	6
7 Gliederungsvorschlag	7
Literaturverzeichnis	9

1 Einleitung

Die Anwendungsmöglichkeiten von KI-Texterzeugungssystemen wie z.B. GPT-3 erstrecken sich mittlerweile über eine Vielzahl von Anwendungsgebieten. Ob im Onlineshop oder im Support als Chatbot der die Kunden berät, in der Softwareentwicklung als Unterstützung für Entwickler [1] bis hin zu Systemen die ganzen Fließtexte, z.B. für Nachrichtenartikel, erstellen können [2]. Durch die immer weiter voranschreitende Entwicklung von neuronalen Netzwerken, auf denen solche Texterzeugungssysteme basieren, wächst das Potential täglich. Eines der Aktuell wohl mächtigsten Modelle ist GPT-3 von OpenAI.

Neben den vielen Anwendungsmöglichkeiten gilt es aber auch die verschiedenen Gefahren und Risiken, die solche Systeme bergen, zu beachten. Während der tatsächlichen Nutzung entsprechender Modelle funktionieren diese als Blackbox. Das bedeutet, dass insbesondere die Daten, mit denen das Modell angelernet wird, extrem wichtig sind. So z.B. der Fall von „Tay“ einem von Microsoft entwickelten Twitter-Bot der aufgrund der an ihn geschickten Nachrichten Lernen und dann selbstständig Twittern sollte. Wegen der vielen negativen Nachrichten von denen „Tay“ lernte, verwandelten sich seine Tweets innerhalb von weniger als 24 Stunden von „[...] humans are super cool“ zu „Hitler was right I hate the jews.“. Sodass Microsoft „Tay“ nach kürzester Zeit wieder offline schaltete [3].

Diese Texterzeugungssysteme zu verstehen ist aufgrund der Komplexität von neuronalen Netzwerken nicht ganz einfach. Ich denke allerdings, dass es aufgrund der Vielzahl an Möglichkeiten, die solche Systeme bieten, immer wichtiger ist, diese zu verstehen. Denn nur wer ein zumindest grundlegendes Verständnis von der Funktionsweise hat, kann auch deren Möglichkeiten und Risiken verstehen und erklären. Nur so ist es möglich, das volle Potential von Texterzeugungssystemen auszunutzen und dabei die vielen Probleme zu umgehen.

Deshalb möchte ich mich in meiner Bachelorarbeit mit dem Thema „Texterzeugungssystemen“ befassen. In diesem Exposé soll nun erst einmal ein Überblick über die Bachelorarbeit gegeben werden. Es werden wichtige Grundlagen kurz erklärt, das Forschungsziel und die Forschungsmethodik erläutert sowie ein Titelvorschlag gemacht. Zudem folgt eine erste Literaturrecherche. Am Ende des Exposés wird zudem ein erster Gliederungsvorschlag für die Bachelorarbeit sowie ein Zeitplan erstellt.

2 Grundlagen

Im Kapitel Grundlagen werden einige Begriffe kurz erklärt, die für das Verständnis des Themas entscheidend sind. In der Bachelorarbeit werden diese dann umfangreicher erläutert.

2.1 Neuronale Netze/Deep Learning

Im Bereich der KI sind mit neuronalen Netzen (NN) in erster Linie künstliche neuronale Netze (KNN) gemeint. Diese basieren grob auf der Funktionsweise eines echten Biologischen NN wie es z.B. im Gehirn vorkommt. Dabei gibt es beim KNN eine Vielzahl an „Neuronen“ die jeweils verschiedene Gewichtungen und Schwellwerte haben. Durch ihre Verknüpfung kann ein KNN Modell dann bestimmte Entscheidungen treffen. [4]

2.2 Texterzeugungssysteme

Ein KNN kann für viele unterschiedliche Aufgaben trainiert werden, ein Texterzeugungssystem ist ein KNN Modell, das auf die Erzeugung von Text trainiert wurde. Dabei ist die Art des zu erzeugenden Textes einzig von den Trainingsdaten abhängig. Wie bereits in der Einleitung beschrieben, ist es möglich, mittels eines solchen Modells, alles, von langen Fließtexten bis hin zu Programmcode, zu erzeugen.

2.3 GPT

GPT steht für „Generative Pre-trained Transformer“. Dabei handelt es sich um eine Reihe von Natural Language Processing- bzw. Texterzeugungsmodellen deren neuste Version, GPT-3, im Mai 2020 von OpenAI veröffentlicht wurde. GPT-1 und GPT-2 stehen dabei der Öffentlichkeit zur Verfügung. Der Zugriff auf GPT-3 ist allerdings erstmalig kostenpflichtig und zurzeit nur ausgewählten Personen vorbehalten. Bei der GPT Reihe, insbesondere bei GPT-3, handelt es sich um eines der aktuell umfangreichsten und mächtigsten Texterzeugungssysteme. [5]

2.4 OpenAI

OpenAI ist eine gemeinnützige Organisation, die 2015 von Elon Musk sowie dem Programmierer Sam Altman und weiteren Investoren gegründet wurde. OpenAI erforscht und entwickelt verschiedene KI-Modelle, so auch die GPT Reihe. [5]

3 Forschungsziel

Der Fokus der Bachelorarbeit liegt in der Auseinandersetzung mit Texterzeugungssystemen. Dabei geht es auf der einen Seite um die Funktionsweise eines solchen Systems. Ziel ist es ein tiefgründiges Verständnis für die Funktionsweise von neuronalen Netzen im Allgemeinen und Texterzeugungssystemen im Speziellen aufzubauen. Dabei geht es sowohl um ein theoretisches, aber auch um ein praktisches Verständnis dieser Systeme.

Auf der anderen Seite stehen die Möglichkeiten und Risiken im Mittelpunkt. Nachdem ein Verständnis für die Technik existiert, ist es wichtig zu wissen, welche Anwendungsmöglichkeiten es gibt. Dabei sollen verschiedene Anwendungsmöglichkeiten diskutiert werden und gegebenenfalls sogar gewisse Geschäftsideen kurz umrissen werden. Bei diesen ist es insbesondere wichtig, die theoretischen Anwendungsideen mit dem aktuellen Stand der Technik zu vergleichen, um so auch die praktische Relevanz von verschiedenen Anwendungsmöglichkeiten einordnen zu können. Mindestens genauso wichtig wie das Erörtern der möglichen Anwendungen ist es, auch die Risiken, die mit Texterzeugungssystemen verbunden sind, herauszustellen. Dabei sind auf der einen Seite die Risiken in Bezug auf spezielle Anwendungsfälle zu beachten und auf der anderen Seite allgemeine Risiken solcher Systeme, die über alle Anwendungsfälle hinweg existieren.

Alles in allem ist es also das Ziel Texterzeugungssysteme in ihrer Gesamtheit zu verstehen. Also über die Funktionsweise, Möglichkeiten und Risiken informiert zu sein und dabei auch eine eigene kritische Einschätzung der Möglichkeiten und Risiken zu geben.

4 Forschungsmethodik

Im ersten Teil der Arbeit liegt der Fokus auf der Recherche, um ein Verständnis für die Funktionsweise von Texterzeugungssystemen herzustellen. Hier wird in erster Linie eine umfangreiche Literaturrecherche stattfinden. Um allerdings auch im ersten Teil den Bezug zur Praxis nicht zu verlieren, sollen hier auch einige praktische Versuche stattfinden, dazu wird die GPT-Reihe verwendet. Da GPT-3 nur schwer zugänglich ist, wird hier vermutlich in erster Linie GPT-2 exemplarisch für Texterzeugungssysteme verwendet.

Im anderen Teil der Arbeit wird etwas Praxisnäher gearbeitet. Hier liegt der Fokus weniger auf Recherche und mehr auf der eigenständigen Erarbeitung von verschiedenen Anwendungs-/Geschäftsideen sowie deren Risiken. Natürlich darf aber auch in diesem Teil die Recherche nicht fehlen, um die Möglichkeiten und Risiken mit Fakten zu untermauern.

5 Titelvorschlag

Funktionsweise, Anwendungen und Gefahren von Texterzeugungsmodellen am Beispiel von GPT

6 Literaturrecherche

- [6] In diesem Artikel wird GPT-3 kurz erklärt und einige Tests mit GPT-3 durchgeführt. Insbesondere werden aber auch einige philosophische Aspekte die für die Möglichkeiten und Risiken von GPT-3 interessant sind angesprochen. Es ist allerdings zu beachten das dieser Artikel, aufgrund seines Themas, sehr subjektiv ist.
- [4] In diesem Buch wird ein Überblick über KI im generellen gegeben. Dabei wird die funktionsweise von neuronalen Netzen kurz erklärt und auch auf Dinge wie Hardware und Normen im Bereich KI eingegangen. Auch Anwendungsmöglichkeiten von KI-Systemen werden diskutiert.
- [7] Neben einer weiteren Erklärung der Funktionsweise von KI-Systemen werden hier vor allem viele Anwendungsfälle von KI im Allgemeinen diskutiert. Zwar geht es dabei nicht speziell um Texterzeugungsmodelle es können aber verschiedene Ideen adaptiert werden.
- [8] Auch wenn der Titel des Buches im ersten Moment etwas reißerisch klingt, bietet das Buch einige interessante Abschnitte. Besonders wichtig ist dabei das Kapitel „Computer lernen sprechen“ da es dabei speziell um KI-Systeme in Bezug auf Sprache/Texte geht, was eher selten zu finden ist.
- [9] Insbesondere das Kapitel 5.3 „Neuronale Netze“ ist hier interessant. Es bietet eine etwas tiefgründigere, aber dennoch anschauliche Erklärung verschiedener Konzepte eines neuronalen Netzes. So wird zum Beispiel Backpropagation erklärt.

7 Gliederungsvorschlag

- 1 Einleitung
 - 1.1 Motivation
 - 1.2 Ablauf
- 2 Funktionsweise
 - 2.1 Theoretische Erklärung von neuronalen Netzen/Deep Learning
 - 2.2 OpenAI GPT
 - 2.2.1 Einführung
 - 2.2.2 Programmbeispiele mit GPT-2 (oder falls möglich GPT-3)
 - 2.2.3 Erfahrung/Schlussfolgerungen aus den Programmbeispielen
- 3 Anwendungsmöglichkeiten
 - 3.1 Allgemeine Nennung verschiedener Anwendungsbereiche-/Möglichkeiten
 - 3.2 Zwei Anwendungsfälle genauer beschrieben
 - 3.2.1 Anwendungsmöglichkeit 1
 - 3.2.2 Anwendungsmöglichkeit 2
 - 3.2.3 Vergleich beider Fälle mit eigener Bewertung
- 4 Risiken und Probleme
 - 4.1 Allgemeine Nennung verschiedener Risiken und Probleme
 - 4.2 Risiken und Probleme der beiden Anwendungsmöglichkeiten
 - 4.2.1 Anwendungsmöglichkeit 1
 - 4.2.2 Anwendungsmöglichkeit 2
 - 4.2.3 Vergleich beider Fälle mit eigener Bewertung
- 5 Ausblick und Fazit

8 Zeitplan

Für die Bachelorarbeit sind eigentlich 9 Wochen vorgesehen. In meinem Zeitplan arbeite ich allerdings mit 8 Wochen um 1 Woche als Zeitpuffer übrig zu haben, falls einer der Schritte länger dauert als geplant.

Woche 1: Einarbeitung ins Thema mit tieferer Literaturrecherche

Woche 2: Schreiben des Kapitels 2.1 (Theoretische Erklärung) und Anfänge für 2.2 (gpt)

Woche 3: Arbeiten mit GPT

Woche 4-5: erarbeiten der Anwendungsmöglichkeiten

Woche 6-7: erarbeiten der Risiken und Probleme

Woche 8: Fehlerkorrektur und gegebenenfalls Überarbeitung einzelner Teile

Literaturverzeichnis

- [1] M. Schreiner, „Mixed,“ [Online]. Available: <https://mixed.de/openai-gpt-3-ki-verwandelt-sprache-zu-code/>. [Zugriff am 31 05 2021].
- [2] GPT-3, „The Guardian,“ [Online]. Available: <https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3>. [Zugriff am 31 05 2021].
- [3] A. Steiner, „faz,“ [Online]. Available: <https://www.faz.net/aktuell/wirtschaft/netzwirtschaft/microsofts-bot-tay-wird-durch-nutzer-zum-nazi-und-sexist-14144019.html>. [Zugriff am 31 05 2021].
- [4] V. Wittpahl, „Künstliche Intelligenz,“ in *Künstliche Intelligenz*, Berlin, Heidelberg, Springer Vieweg, 2019, pp. 31-32.
- [5] T. Nord, „Lernen Wie Maschinen,“ 19 04 2021. [Online]. Available: <https://www.lernen-wie-maschinen.ai/ki-pedia/was-ist-gpt-3-und-spricht-das-modell-deutsch/>. [Zugriff am 01 06 2021].
- [6] L. Floridi und M. Chiriatti, „GPT-3: Its Nature, Scope, Limits, and Consequences,“ *Minds & Machines*, Nr. 30, p. 681–694, 2020.
- [7] P. Wennker, *Künstliche Intelligenz in der Praxis*, Wiesbaden : Springer Fachmedien Wiesbaden GmbH, 2020.
- [8] K. Mainzer, *Künstliche Intelligenz – Wann übernehmen die Maschinen?*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2019.
- [9] M. von Rimscha, *Algorithmen kompakt und verständlich*, Wiesbaden: Springer Fachmedien Wiesbaden, 2014.