

# Aufgabenblatt 01

Abgabe: 17.11, 11:59 Uhr

Gesamtpunkte: 50

Notwendig zum Bestehen: 25

## Aufgabe 1 – Language Modelle und ihr aktueller Diskurs (10 Punkte)

Dies ist eine wiederkehrende Aufgabe auf jedem der ersten fünf Aufgabenblätter. Ziel ist es, den aktuellen Diskurs rund um Large Language Models besser zu durchdringen und in einem wissenschaftlich-kritischen Kontext einzuordnen.

- 1.1 Wählt einen aktuellen Beitrag aus (nicht älter als 4 Wochen), bei dem Large Language Modelle eine zentrale Rolle spielen. Der Beitrag kann entweder technisch die Implementierung oder neue Feature von LLMs beschreiben, eine Anwendung von Language-Modellen vorstellen, oder auch eine sonstige rechtliche, wirtschaftliche oder soziale Einordnung diskutieren. Der Beitrag kann in einem Blog, einer Nachrichtenseite oder in den sozialen Medien veröffentlicht worden sein.  
Gebt den Titel, Autor, Medium der Veröffentlichung, Datum der Veröffentlichung sowie einen Link zum Beitrag an.  
Beschreibt in ein paar kurzen Sätzen, warum ihr den Beitrag ausgewählt habt.  
(2 Punkte)
- 1.2 Fasst den Inhalt des Beitrags zunächst kurz mit euren eigenen Worten zusammen.  
(ca. 100 Worte, 2 Punkte)
- 1.3 Analyse und kritische Reflexion: Ordnet den Beitrag nun kritisch ein. Wenn eine technische Neuerung vorgestellt wird, dann erörtert die neuen Möglichkeiten, aber auch welche kritischen Folgen sich daraus ergeben können. Handelt es sich um eine Anwendung, diskutiert auch hier kritisch, welche Probleme dadurch gelöst werden, aber auch welche neuen Probleme entstehen können. Ihr könnt bei der Einordnung gern weitere Quellen heranziehen. Vielleicht haben bereits andere ebenfalls darüber berichtet? Endet mit einem eigenen Fazit, in dem ihr auch erörtert, welche Fragen noch offenbleiben. (ca. 500 – 1000 Worte, 6 Punkte)

## Aufgabe 2 – Datensatzanalyse (20 Punkte)

Für diese und die kommende Aufgabe beschäftigen wir uns mit einem Datensatz von mdr Nachrichten in leichter Sprache. Den Datensatz findet ihr auf Moodle.

Sucht euch für diese Aufgabe zunächst fünf Artikel aus dem Datensatz aus und findet die entsprechenden Artikel auch in normaler / schwerer Sprache. Dazu könnt ihr das Internet-Archiv unter <https://web.archive.org/> benutzen.

Ziel der Aufgabe ist es zu verstehen, inwieweit sich die Daten aus dem Datensatz der leichten Sprache von gewöhnlichen Datensätzen unterscheiden.

- 2.1 Sucht euch zunächst zwei Metriken aus, bei denen ihr erwartet, dass sich eure Daten vom Datensatz in leichter Sprache nicht von der regulären Sprache unterscheiden. Erläutert für beide Metriken, warum ihr erwartet, dass sich diese nicht unterscheiden. (4 Punkte)
- 2.2 Sucht euch nun zwei Metriken aus, bei denen ihr einen Unterschied in den beiden Datensätzen erwartet. Erläutert wieder, warum ihr hier einen Unterschied in den Datensätzen erwartet. (4 Punkte)
- 2.3 Implementiert nun eure Metriken und analysiert die ausgewählten Artikel entsprechend. Erläutert dann, inwieweit sich eure Hypothesen bewahrheitet haben. (12 Punkte)  
Gebt für diese Aufgabe auch euren Programmcode ab.

## Aufgabe 4 – Bigram-Modell zur Artikelgenerierung (20 Punkte)

In dieser Aufgabe wollen wir jetzt, ähnlich wie in der Vorlesung, ein ganz einfaches Modell erstellen, dass für jedes Wort immer das nächstwahrscheinliche vorhersagt und so ganze Nachrichten-Artikel generiert. Nutzt hierfür nur den Datensatz in leichter Sprache.

3.1 Wandelt die Text in einen Trainingsdatensatz um, den ihr zum Training eines simplen neuronalen Netzes verwenden könnt. (5 Punkte)

3.2 Implementiert nun wie in der Vorlesung ein simples Neuronales Netz, das als Input-Layer pro möglichem Wort in eurem Input-Datensatz ein Neuron und genau so pro möglichem Wort ein Neuron im Output Layer besitzt. Initialisiert die Gewichtsmatrix zufällig und lasst das Netz nun trainieren. Experimentiert dabei sowohl mit verschiedenen Lernraten als auch mit einer verschiedenen großen Anzahl von Update-Zyklen. Was fällt euch dabei auf, welche Kombination scheint gut zu funktionieren? Erläutert kurz. (10 Punkte)

3.3 Generiert nun mit Hilfe eures Neuronalen Netzes einen einfachen Text von ca. 100 Worten. Was haltet ihr vom Ergebnis des Texts? Wie nah erscheint er euch als Leser:in an den tatsächlichen Texten im Trainingsdatensatz zu sein? Woher kommt eine mögliche Abweichung? Erläutert kurz. (5 Punkte)

Gebt zu Aufgabe 3.1 und 3.2 auch euren Programmcode ab. Für 3.3, gebt auch den von euch generierten Text ab.

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan et iusto odio dignissim qui blandit praesent luptatum zzril delenit augue duis dolore te feugait nulla facilisi. Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat.

Ut wisi enim ad minim veniam, quis nostrud exerci tation ullamcorper suscipit lobortis nisl ut aliquip ex ea commodo consequat. Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan et iusto odio dignissim qui blandit praesent luptatum zzril delenit augue duis dolore te feugait nulla facilisi.

Nam liber tempor cum soluta nobis eleifend option congue nihil imperdiet doming id quod mazim placerat facer possim assum. Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat. Ut wisi enim ad minim veniam, quis nostrud exerci tation ullamcorper suscipit lobortis nisl ut aliquip ex ea commodo consequat.

Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis.

At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, At accusam aliquyam diam diam dolore

dolores duo eirmod eos erat, et nonumy sed tempor et et invidunt justo labore Stet clita ea et gubergren, kasd magna no rebum. sanctus sea sed takimata ut vero voluptua. est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat.

Consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus.

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan et iusto odio dignissim qui blandit praesent luptatum zzril delenit augue duis dolore te feugait nulla facilisi. Lorem ipsum dolor sit amet, consetetur adipiscing elit, sed diam nonumy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat.

Ut wisi enim ad minim veniam, quis nostrud exerci tation ullamcorper suscipit lobortis nisl ut aliquip ex ea commodo consequat. Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan et iusto odio dignissim qui blandit praesent luptatum zzril delenit augue duis dolore te feugait nulla facilisi.

Nam liber tempor cum soluta nobis eleifend option congue nihil imperdiet doming id quod mazim placerat facer possim assum. Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat. Ut wisi enim ad minim veniam, quis nostrud exerci tation ullamcorper suscipit lobortis nisl ut aliquip ex ea commodo consequat.

Duis autem vel eum iriure dolor in