

# **Detaillierter Projektstrukturplan**

## **Dokumentation des Moduls "Projektrealisierung"**

im Bereich Wirtschaftsinformatik: Data Science  
an der Dualen Hochschule Baden-Württemberg Mannheim

Autoren: Niklas Koch, Niclas Cramer, Jasmina Pascanovic & Antoine Fuchs  
Matrikelnummern: 6699912, 7607733, 5711726 & 3008441  
Kurs: WWI20DSA  
Studiengangsleiter: Prof. Dr.-Ing. habil. Dennis Pfisterer  
Dozent: Michael Lang, Enzo Hilzinger  
Abgabedatum: 20.06.2023

# Inhaltsverzeichnis

<b>Inhaltsverzeichnis</b>	<b>2</b>
<b>1 Projektplanung</b>	<b>5</b>
1.1 Planung & Management . . . . .	5
1.1.1 Ressourcenmanagement . . . . .	5
1.1.2 Stakeholderanalyse . . . . .	6
1.1.3 Zeitplanung . . . . .	7
1.1.4 Vorbereitung der Meetings mit den Dozenten . . . . .	8
1.2 Erstellung von Abgaben . . . . .	9
1.2.1 Erstellung der Präsentation 1 . . . . .	9
1.2.2 Erstellung der Präsentation 2 . . . . .	10
1.2.3 Erstellung der Präsentation 3 . . . . .	11
1.2.4 Erstellung des Projektauftrags . . . . .	12
1.2.5 Erstellung des Lasten- und Pflichtenhefts . . . . .	13
1.2.6 Erstellung des groben Projektstrukturplans . . . . .	14
1.2.7 Erstellung des detaillierten Projektstrukturplans . . . . .	15
1.2.8 Erstellung des Gantt-Charts . . . . .	16
1.2.9 Erstellung des Projektabschlussberichts . . . . .	17
1.2.10 Abgabe der ersten Prüfungsleistung . . . . .	18
1.2.11 Abgabe der zweiten Prüfungsleistung . . . . .	19
1.2.12 Abgabe des Projekts zur Bewertung . . . . .	20
<b>2 Datenbeschaffung</b>	<b>21</b>
2.1 Finden geeigneter Daten . . . . .	21
2.1.1 Finden geeigneter Daten für „Wissenschaftliche Texte“ . . . . .	21
2.1.2 Finden geeigneter Daten für „Nachrichtenartikel“ . . . . .	22
2.1.3 Finden geeigneter Daten für „Reviews“ . . . . .	23
2.1.4 Finden geeigneter Daten für „Literarische Texte“ . . . . .	24
2.2 Verarbeitung der Daten . . . . .	25
2.2.1 Vorverarbeiten & Zusammenfassen zu einem Gesamtdatensatz . . . . .	25
2.2.2 Feature-Extraktion der Texte . . . . .	26
2.2.3 Daten zur Verfügung stellen . . . . .	28
2.3 Visualisierungen . . . . .	29
2.3.1 Grobe Datenanalyse (Data Exploration) . . . . .	29
2.3.2 Visualisierung der zusätzlichen Features . . . . .	30

---

<b>3</b>	<b>Klassifikation</b>	<b>31</b>
3.1	Signifikanztests für zusätzliche Features . . . . .	31
3.2	Umsetzung der Klassifikation mit simplem Algorithmus . . . . .	32
3.3	Umsetzung der Klassifikation mit eigenem neuronalen Netz . . . . .	33
3.4	Optimierung des eigenen neuronalen Netzes zur Klassifikation . . . . .	34
3.5	Umsetzung der Klassifikation mit einem Transformer-Modell . . . . .	35
3.6	Optimierung des Transformer-Modells zur Klassifikation . . . . .	36
3.7	Vergleich der verschiedenen Verfahren auf Effektivität . . . . .	37
<b>4</b>	<b>Zusammenfassung</b>	<b>38</b>
4.1	Gestaltung einer Pipeline . . . . .	38
4.1.1	Recherche einer Trainingspipeline . . . . .	38
4.1.2	Implementierung der Vorverarbeitung . . . . .	40
4.1.3	Erstellen von Referenz-Zusammenfassungen für das Training der Modelle . . . . .	41
4.1.4	Erstellen der Trainingspipeline für die Modellerstellung . . . . .	42
4.1.5	Erstellen der Evaluationspipeline für die Kompressionsrate und den ROUGE-Score . . . . .	43
4.2	Gestaltung von eigens trainierten Modellen . . . . .	45
4.2.1	Recherche von Transformer-Modellen . . . . .	45
4.2.2	Testen der Basis-Transformer-Modelle für die Auswahl eines passenden Transformer-Modells . . . . .	46
4.2.3	Trainieren der Modelle auf Basis der Trainingsdaten und Referenz-Zusammenfassungen . . . . .	47
4.2.4	Evaluierung der verschiedenen Modelle auf Basis des ROUGE-Scores . . . . .	48
4.2.5	Evaluierung der verschiedenen Modelle auf Basis der Kompressionsrate . . . . .	49
<b>5</b>	<b>Entwicklung von zusätzlichen Modulen</b>	<b>50</b>
5.1	Entwicklung einer Spracheingabemöglichkeit . . . . .	50
5.1.1	Recherche von Spracheingabemöglichkeiten . . . . .	50
5.1.2	Identifizieren von geeigneten Kommandos für die Spracheingabennutzung der Anwendung . . . . .	52
5.1.3	Implementierung einer englischen Spracheingabefunktion . . . . .	53
5.1.4	Testen der englischen Spracheingabefunktion . . . . .	54
5.2	Entwicklung einer Möglichkeit Dokumente einzulesen . . . . .	55
5.2.1	Recherche von Möglichkeiten zum Einlesen von Dokumenten . . . . .	55

5.2.2	Implementierung des Einlesens von Dokumenten . . . . .	56
5.2.3	Testen auf verschiedenen Betriebssystemen und mit verschiedenen Dateiformaten . . . . .	58
<b>6</b>	<b>Frontend</b>	<b>60</b>
6.1	Anbindungen . . . . .	60
6.1.1	Anbindung von trainierten Modellen an das Frontend . . . . .	60
6.1.2	Anbindung einer Spracheingabe . . . . .	62
6.1.3	Anbindung der Dokumente-Einlesen-Funktion . . . . .	64
6.2	Ausgabe . . . . .	66
6.2.1	Ausgabe des Klassifikationsergebnisses im Frontend . . . . .	66
6.2.2	Ausgabe des Zusammenfassungsergebnisses im Frontend . . . . .	68
6.2.3	Integration der Pipeline zur Durchführung der Klassifikation und Zu- sammenfassung im Frontend . . . . .	70
6.3	Barrierefreiheit . . . . .	72
6.3.1	Recherche für ein barrierefreies Frontend . . . . .	72
6.3.2	Design eines barrierefreien Frontends . . . . .	73
6.3.3	Umsetzung und Implementierung des barrierefreien Frontend-Designs	75
6.4	Testen der gesamten Frontend-Anwendung . . . . .	77
<b>7</b>	<b>Dokumentation</b>	<b>79</b>
7.1	Dokumentation der allgemeinen Aspekte . . . . .	79
7.1.1	Dokumentation der Datenbeschaffung . . . . .	79
7.1.2	Dokumentation des Moduls zur Spracheingabe . . . . .	80
7.1.3	Dokumentation durch Kommentieren des Codes zur Spracheingabe	81
7.1.4	Dokumentation des Moduls zum Einlesen von Textdokumenten . .	82
7.1.5	Dokumentation durch Kommentieren des Codes zum Einlesen von Textdokumenten . . . . .	83
7.2	Dokumentation der Klassifikation . . . . .	84
7.2.1	Dokumentation der einzelnen Module zur Klassifikation . . . . .	84
7.2.2	Dokumentation durch Kommentieren des Codes der Klassifikation .	85
7.3	Dokumentation der Zusammenfassung . . . . .	86
7.3.1	Dokumentation der einzelnen Module zur Zusammenfassung . . .	86
7.3.2	Dokumentation durch Kommentieren des Codes zur Zusammenfas- sung . . . . .	87

# 1 Projektplanung

## 1.1 Planung & Management

### 1.1.1 Ressourcenmanagement

**PSP-Code:** AP-111

**Verantwortlicher:** Jasmina Pascanovic

**Aufgabenstellung:** Das Ressourcenmanagement soll durchgeführt werden, indem eine Akkumulation der verfügbaren Ressourcen Zeit, Rechenleistung und Personal erfolgt, um diese daraufhin auf spezifische, zu erledigende Arbeitspakete zu verteilen.

**Zieldefinition:** Die Ressourcen müssen klar definiert und auf den Projektzeitraum passend verteilt werden.

**Geplanter Zeitraum:** 08.05.23 - 31.05.23

**Voraussetzung:** Die Anforderungen an das Projekt werden vorausgesetzt, um die Einteilung durchführen zu können.

**Hilfsmittel:** LaTeX, Discord

**Risiken:** Ein unzureichendes Ressourcenmanagement kann zu einer Verzögerung des Projektfortschritts führen, wodurch die Einhaltung von Meilensteinen und Teilzielen erschwert werden kann.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Dieses Arbeitspaket ist grundlegend für alle darauf aufbauenden, da mit den wichtigsten Ressourcen geplant wird, allerdings müssen die Arbeitspakete hier bereits grob erstellt worden sein.

**Abnahmebedingungen:** Das Ressourcenmanagement wurde ausführlich durchgeführt und mit dem Projektteam abgestimmt.

### 1.1.2 Stakeholderanalyse

**PSP-Code:** AP-112

**Verantwortlicher:** Jasmina Pascanovic

**Aufgabenstellung:** Die Stakeholderanalyse soll durchgeführt werden, indem vor allem die Erwartungen und Haltungen zu dem Projektumfang klar kommuniziert werden. Dazu fungieren in diesem Fall auch die Dozenten als Stakeholder.

**Zieldefinition:** Die Stakeholder müssen identifiziert und eingeordnet werden.

**Geplanter Zeitraum:** 08.05.23 - 15.05.23

**Voraussetzung:** Die Anforderungen an das Projekt werden vorausgesetzt, um die Einteilung durchführen zu können.

**Hilfsmittel:** LaTeX, Discord, Microsoft Teams

**Risiken:** Eine unzureichende Stakeholderanalyse kann durch Unklarheiten zu einer Verzögerung des Projektfortschritts führen, wodurch die Einhaltung von Meilensteinen und Teilzielen erschwert werden kann.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Dieses Arbeitspaket ist grundlegend für alle darauf aufbauenden.

**Abnahmebedingungen:** Die Stakeholderanalyse wurde ausführlich durchgeführt und mit dem Projektteam abgestimmt.

### 1.1.3 Zeitplanung

**PSP-Code:** AP-113

**Verantwortlicher:** Jasmina Pascanovic

**Aufgabenstellung:** Die Zeitplanung soll durchgeführt werden, indem eine Klarstellung der Meilensteine und Teilziele erfolgt.

**Zieldefinition:** Die zeitlichen Ressourcen müssen klar definiert und auf den Projektzeitraum passend verteilt werden.

**Geplanter Zeitraum:** 08.05.23 - 31.05.23

**Voraussetzung:** Die Meilensteine und Teilziele werden aus den Vorgaben der Dozenten und dem bis dato festgelegten Scope des Projekts gefolgert.

**Hilfsmittel:** LaTeX, Discord

**Risiken:** Eine fehlerhafte Zeitplanung kann zu einer Verzögerung des Projektfortschritts führen, wodurch die Einhaltung von Meilensteinen und Teilzielen erschwert werden kann.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Dieses Arbeitspaket ist grundlegend für alle darauf aufbauenden, da mit den wichtigsten Ressourcen geplant wird, allerdings ist die Festlegung des Scopes vorher notwendig.

**Abnahmebedingungen:** Die Zeitplanung wurde ausführlich und realistisch durchgeführt sowie mit dem Projektteam abgestimmt.

### **1.1.4 Vorbereitung der Meetings mit den Dozenten**

**PSP-Code:** AP-114

**Verantwortlicher:** Niklas Koch, Niclas Cramer, Jasmina Pascanovic & Antoine Fuchs

**Aufgabenstellung:** Eine Abstimmung für die Meetings mit den Dozenten muss erfolgen, um einen effizienten Austausch in kurzer Zeit zu gewährleisten.

**Zieldefinition:** Etwaige Berichte zum Fortschritt oder unklare Themen müssen innerhalb des Projektteams vor den festgelegten Meetings abgestimmt werden.

**Geplanter Zeitraum:** 08.05.23 - 24.07.23

**Voraussetzung:** Kontinuierlicher Fortschritt über den Projektzeitraum hinweg.

**Hilfsmittel:** Discord

**Risiken:** Eine unzureichende Vorbereitung von den Meetings kann dazu führen, dass die geplante Zeit deutlich überschritten wird oder auch vereinzelt überflüssige Fragen an die Dozenten gestellt werden.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Keine

**Abnahmebedingungen:** Die Vorbereitung für die Meetings ist vor jedem Termin erfolgt.



## **1.2 Erstellung von Abgaben**

### **1.2.1 Erstellung der Präsentation 1**

**PSP-Code:** AP-121

**Verantwortlicher:** Niklas Koch, Niclas Cramer, Jasmina Pascanovic & Antoine Fuchs

**Aufgabenstellung:** Die Erstellung der ersten Präsentation mit einer Länge von 30min muss nach den gegebenen Vorgaben erfolgen.

**Zieldefinition:** Die erste Präsentation wird vollständig ausgearbeitet.

**Geplanter Zeitraum:** 08.05.23 - 24.05.23

**Voraussetzung:** Die Initialisierung des Projekts inklusive aller bis dato festgelegten Anforderungen müssen erledigt sein.

**Hilfsmittel:** PowerPoint, Discord

**Risiken:** Keine

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Dieses Arbeitspaket enthält alle zeitlich vorangegangenen, da diese jeweils in die Präsentation mit einbezogen werden.

**Abnahmebedingungen:** Die Präsentation wurde vollständig ausgestaltet und vorbereitet.

### 1.2.2 Erstellung der Präsentation 2

**PSP-Code:** AP-122

**Verantwortlicher:** Niklas Koch, Niclas Cramer, Jasmina Pascanovic & Antoine Fuchs

**Aufgabenstellung:** Die Erstellung der zweiten Präsentation mit einer Länge von 15min muss nach den gegebenen Vorgaben erfolgen.

**Zieldefinition:** Die zweite Präsentation wird vollständig ausgearbeitet.

**Geplanter Zeitraum:** 24.05.23 - 20.06.23

**Voraussetzung:** Die Arbeitspakete, die zeitlich vor der Präsentation liegen, müssen erledigt sein, damit die Ergebnisse in die Präsentation mit einfließen können.

**Hilfsmittel:** PowerPoint, Discord

**Risiken:** Keine

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Dieses Arbeitspaket enthält alle zeitlich vorangegangenen, da diese jeweils in die Präsentation mit einbezogen werden.

**Abnahmebedingungen:** Die Präsentation wurde vollständig ausgestaltet und vorbereitet.

### 1.2.3 Erstellung der Präsentation 3

**PSP-Code:** AP-123

**Verantwortlicher:** Niklas Koch, Niclas Cramer, Jasmina Pascanovic & Antoine Fuchs

**Aufgabenstellung:** Die Erstellung der dritten Präsentation mit einer Länge von 45min muss nach den gegebenen Vorgaben erfolgen.

**Zieldefinition:** Die dritte Präsentation wird vollständig ausgearbeitet.

**Geplanter Zeitraum:** 20.06.23 - 27.07.23

**Voraussetzung:** Die Arbeitspakete, die zeitlich vor der Präsentation liegen, müssen erledigt sein, damit die Ergebnisse in die Präsentation mit einfließen können.

**Hilfsmittel:** PowerPoint, Discord

**Risiken:** Keine

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Dieses Arbeitspaket enthält alle zeitlich vorangegangenen, da diese jeweils in die Präsentation mit einbezogen werden.

**Abnahmebedingungen:** Die Präsentation wurde vollständig ausgestaltet und vorbereitet.

### 1.2.4 Erstellung des Projektauftrags

**PSP-Code:** AP-124

**Verantwortlicher:** Niklas Koch, Niclas Cramer, Jasmina Pascanovic & Antoine Fuchs

**Aufgabenstellung:** Die Erstellung des Projektauftrags erfolgt nach der Vorlage verschiedener Quellen.

**Zieldefinition:** Der Projektauftrag wird vollständig ausgearbeitet.

**Geplanter Zeitraum:** 08.05.23 - 24.05.23

**Voraussetzung:** Die Anforderungen an das Projekt sind bekannt.

**Hilfsmittel:** LaTeX, Discord

**Risiken:** Ein unvollständiger Projektauftrag kann dazu führen, dass die Abgabe nicht vollständig ist und eine Reduktion der Punktzahl folgt.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Dieses Arbeitspaket ist als Grundlage für die anderen anzusehen.

**Abnahmebedingungen:** Der Projektauftrag wurde vollständig ausgearbeitet.

### 1.2.5 Erstellung des Lasten- und Pflichtenhefts

**PSP-Code:** AP-125

**Verantwortlicher:** Niklas Koch, Niclas Cramer, Jasmina Pascanovic & Antoine Fuchs

**Aufgabenstellung:** Die Erstellung des Lasten- und Pflichtenhefts erfolgt nach der Vorlage verschiedenster Quellen.

**Zieldefinition:** Das Lasten- und Pflichtenheft wird vollständig ausgearbeitet.

**Geplanter Zeitraum:** 08.05.23 - 24.05.23

**Voraussetzung:** Die Anforderungen der Dozenten sowie Ideen für Umsetzung liegen vor.

**Hilfsmittel:** LaTeX, Discord

**Risiken:** Ein unvollständiges Lasten- und Pflichtenheft kann dazu führen, dass die Abgabe nicht vollständig ist und eine Reduktion der Punktzahl folgt.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Dieses Arbeitspaket ist als Grundlage für die anderen anzusehen.

**Abnahmebedingungen:** Das Lasten- und Pflichtenheft wurde vollständig ausgearbeitet.

### 1.2.6 Erstellung des groben Projektstrukturplans

**PSP-Code:** AP-126

**Verantwortlicher:** Jasmina Pascanovic

**Aufgabenstellung:** Die Erstellung des groben Projektstrukturplans erfolgt nach der Vorlage verschiedenster Quellen.

**Zieldefinition:** Der grobe Projektstrukturplan wird vollständig ausgearbeitet.

**Geplanter Zeitraum:** 08.05.23 - 24.05.23

**Voraussetzung:** Die Anforderungen der Dozenten sowie Ideen für Umsetzung liegen vor.

**Hilfsmittel:** LaTeX, Discord

**Risiken:** Ein unvollständiger grober Projektstrukturplan kann dazu führen, dass die Abgabe nicht vollständig ist und eine Reduktion der Punktzahl folgt.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Dieses Arbeitspaket ist als Grundlage für die chronologische Abarbeitung der Arbeitspakete, sowie die Zeitplanung, anzusehen.

**Abnahmebedingungen:** Der grobe Projektstrukturplan wurde vollständig ausgearbeitet.

### 1.2.7 Erstellung des detaillierten Projektstrukturplans

**PSP-Code:** AP-127

**Verantwortlicher:** Niklas Koch, Niclas Cramer, Jasmina Pascanovic & Antoine Fuchs

**Aufgabenstellung:** Die Erstellung des detaillierten Projektstrukturplans erfolgt nach der Vorlage verschiedenster Quellen.

**Zieldefinition:** Der detaillierte Projektstrukturplan wird vollständig ausgearbeitet.

**Geplanter Zeitraum:** 24.05.23 - 20.06.23

**Voraussetzung:** Die Anforderungen der Dozenten sowie Ideen für Umsetzung liegen vor.

**Hilfsmittel:** LaTeX, Discord

**Risiken:** Ein unvollständiger detaillierter Projektstrukturplan kann dazu führen, dass die Abgabe nicht vollständig ist und eine Reduktion der Punktzahl folgt.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Dieses Arbeitspaket ist als Grundlage für die chronologische Abarbeitung der Arbeitspakete, sowie die Zeitplanung, anzusehen.

**Abnahmebedingungen:** Der detaillierte Projektstrukturplan wurde vollständig ausgearbeitet.

### 1.2.8 Erstellung des Gantt-Charts

**PSP-Code:** AP-128

**Verantwortlicher:** Jasmina Pascanovic

**Aufgabenstellung:** Die Erstellung des Gantt-Charts wird anhand der Oberkategorien der Arbeitspakete erfolgen.

**Zieldefinition:** Das Gantt-Chart wird vollständig ausgearbeitet.

**Geplanter Zeitraum:** 24.05.23 - 20.06.23

**Voraussetzung:** Der detaillierte Projektstrukturplan liegt vor.

**Hilfsmittel:** Excel, Discord

**Risiken:** Ein unvollständiges Gantt-Chart kann dazu führen, dass die Abgabe nicht vollständig ist und eine Reduktion der Punktzahl erfolgt.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Dieses Arbeitspaket ist als Grundlage für die chronologische Abarbeitung der Arbeitspakete, sowie die intensivere Zeitplanung, anzusehen.

**Abnahmebedingungen:** Das Gantt-Chart wurde vollständig ausgearbeitet.



### 1.2.9 Erstellung des Projektabschlussberichts

**PSP-Code:** AP-129

**Verantwortlicher:** Niklas Koch, Niclas Cramer, Jasmina Pascanovic & Antoine Fuchs

**Aufgabenstellung:** Die Erstellung des Projektabschlussberichts mit den Lessons Learned, einer Reflexion, erfüllten/nicht erfüllten Kriterien aus dem Pflichtenheft und einem Ausblick erfolgt.

**Zieldefinition:** Der Projektabschlussbericht wird vollständig ausgearbeitet.

**Geplanter Zeitraum:** 20.06.23 - 27.07.23

**Voraussetzung:** Die Arbeitspakete wurden möglichst vollständig und nach Zeitplan erledigt.

**Hilfsmittel:** PowerPoint, Discord

**Risiken:** Ein unvollständiger Projektabschlussbericht kann dazu führen, dass die Abgabe nicht vollständig ist und eine Reduktion der Punktzahl erfolgt.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Dieses Arbeitspaket enthält alle zeitlich vorangegangenen, da diese jeweils in den Abschlussbericht mit einbezogen werden.

**Abnahmebedingungen:** Der Projektabschlussbericht wurde vollständig ausgearbeitet.

### 1.2.10 Abgabe der ersten Prüfungsleistung

**PSP-Code:** AP-1210

**Verantwortlicher:** Niklas Koch, Niclas Cramer, Jasmina Pascanovic & Antoine Fuchs

**Aufgabenstellung:** Die Abgabe der ersten Prüfungsleistung, inklusive der ersten Präsentation, des Projektauftrags, des Lasten- und Pflichtenhefts und des groben Projektstrukturplans, erfolgt.

**Zieldefinition:** Die erste Prüfungsleistung wird erfolgreich abgegeben.

**Geplanter Zeitraum:** 08.05.23 - 24.05.23

**Voraussetzung:** Die Arbeitspakete bis dato sollten bestmöglichst erledigt worden sein.

**Hilfsmittel:** LaTeX, GitHub, VSCode, Discord

**Risiken:** Eine unvollständige oder verspätete Abgabe der ersten Prüfungsleistung kann dazu führen, dass die Abgabe nicht vollständig ist und eine Reduktion der Punktzahl erfolgt.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Dieses Arbeitspaket enthält alle zeitlich vorangegangenen, da diese jeweils mit einbezogen werden.

**Abnahmebedingungen:** Die Abgabe wurde bestmöglich, vollständig und in der gegebenen Zeit erledigt.

### 1.2.11 Abgabe der zweiten Prüfungsleistung

**PSP-Code:** AP-1211

**Verantwortlicher:** Niklas Koch, Niclas Cramer, Jasmina Pascanovic & Antoine Fuchs

**Aufgabenstellung:** Die Abgabe der zweiten Prüfungsleistung, inklusive der zweiten Präsentation, des detaillierten Projektstrukturplans und des Gantt-Charts, erfolgt.

**Zieldefinition:** Die zweite Prüfungsleistung wird erfolgreich abgegeben.

**Geplanter Zeitraum:** 24.05.23 - 20.06.23

**Voraussetzung:** Die Arbeitspakete bis dato sollten bestmöglichst erledigt worden sein.

**Hilfsmittel:** LaTeX, GitHub, VSCode, Discord

**Risiken:** Eine unvollständige oder verspätete Abgabe der zweiten Prüfungsleistung kann dazu führen, dass die Abgabe nicht vollständig ist und eine Reduktion der Punktzahl erfolgt.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Dieses Arbeitspaket enthält alle zeitlich vorangegangenen, da diese jeweils mit einbezogen werden.

**Abnahmebedingungen:** Die Abgabe wurde bestmöglich, vollständig und in der gegebenen Zeit erledigt.

### 1.2.12 Abgabe des Projekts zur Bewertung

**PSP-Code:** AP-1212

**Verantwortlicher:** Niklas Koch, Niclas Cramer, Jasmina Pascanovic & Antoine Fuchs

**Aufgabenstellung:** Das gesamte Projekt, inklusive des Projektabschlussberichts und des kommentierten Codes aus dem GitHub Repository, wird abgegeben.

**Zieldefinition:** Die Abgabe des Projekts wird erfolgreich ausgeführt.

**Geplanter Zeitraum:** 20.06.23 - 27.07.23

**Voraussetzung:** Die Arbeitspakete sollten bestmöglichst erledigt worden sein.

**Hilfsmittel:** LaTeX, GitHub, VSCode, Discord

**Risiken:** Eine unvollständige oder verspätete Abgabe des Projekts kann dazu führen, dass die Abgabe nicht vollständig ist und eine Reduktion der Punktzahl erfolgt.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Dieses Arbeitspaket enthält alle zeitlich vorangegangenen, da diese jeweils mit einbezogen werden.

**Abnahmebedingungen:** Die Abgabe wurde bestmöglich, vollständig und in der gegebenen Zeit erledigt.

## 2 Datenbeschaffung

### 2.1 Finden geeigneter Daten

#### 2.1.1 Finden geeigneter Daten für „Wissenschaftliche Texte“

**PSP-Code:** AP-211

**Verantwortlicher:** Niklas Koch

**Aufgabenstellung:** Es müssen geeignete Texte der Kategorie „Wissenschaftliche Texte“ gefunden werden, welche für die Klassifikation und die Zusammenfassung genutzt werden können.

**Zieldefinition:** Ein geeigneter Datensatz mit wissenschaftlichen Texten muss gefunden und heruntergeladen werden.

**Geplanter Zeitraum:** 24.05.23 - 07.06.23.

**Voraussetzung:** Keine.

**Hilfsmittel:** Internet, Kaggle.

**Risiken:** Daten sind ungeeignet.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Die Daten sind die Grundlage des gesamten Projekts und alle anderen Arbeitspakete sind abhängig von diesen.

**Abnahmebedingungen:** Wissenschaftliche Texte sind gefunden und heruntergeladen.

### 2.1.2 Finden geeigneter Daten für „Nachrichtenartikel“

**PSP-Code:** AP-212

**Verantwortlicher:** Niklas Koch

**Aufgabenstellung:** Es müssen geeignete Texte der Kategorie „Nachrichtenartikel“ gefunden werden, welche für die Klassifikation und die Zusammenfassung genutzt werden können.

**Zieldefinition:** Ein geeigneter Datensatz mit Nachrichtenartikeln muss gefunden und heruntergeladen werden.

**Geplanter Zeitraum:** 24.05.23 - 07.06.23.

**Voraussetzung:** Keine.

**Hilfsmittel:** Internet, Kaggle.

**Risiken:** Daten sind ungeeignet.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Die Daten sind die Grundlage des gesamten Projekts und alle anderen Arbeitspakete sind abhängig von diesen.

**Abnahmebedingungen:** Nachrichtenartikel sind gefunden und heruntergeladen.

### 2.1.3 Finden geeigneter Daten für „Reviews“

**PSP-Code:** AP-213

**Verantwortlicher:** Niklas Koch

**Aufgabenstellung:** Es müssen geeignete Texte der Kategorie „Reviews“ gefunden werden, welche für die Klassifikation und die Zusammenfassung genutzt werden können.

**Zieldefinition:** Ein geeigneter Datensatz mit Reviews muss gefunden und heruntergeladen werden.

**Geplanter Zeitraum:** 24.05.23 - 07.06.23.

**Voraussetzung:** Keine.

**Hilfsmittel:** Internet, Kaggle.

**Risiken:** Daten sind ungeeignet.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Die Daten sind die Grundlage des gesamten Projekts und alle anderen Arbeitspakete sind abhängig von diesen.

**Abnahmebedingungen:** Reviews sind gefunden und heruntergeladen.

### 2.1.4 Finden geeigneter Daten für „Literarische Texte“

**PSP-Code:** AP-214

**Verantwortlicher:** Niklas Koch

**Aufgabenstellung:** Es müssen geeignete Texte der Kategorie „Literarische Texte“ gefunden werden, welche für die Klassifikation und die Zusammenfassung genutzt werden können.

**Zieldefinition:** Ein geeigneter Datensatz mit literarischen Texte muss gefunden und heruntergeladen werden.

**Geplanter Zeitraum:** 24.05.23 - 07.06.23.

**Voraussetzung:** Keine.

**Hilfsmittel:** Internet, Kaggle.

**Risiken:** Daten sind ungeeignet.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Die Daten sind die Grundlage des gesamten Projekts und alle anderen Arbeitspakete sind abhängig von diesen.

**Abnahmebedingungen:** Literarische Texte sind gefunden und heruntergeladen.



## 2.2 Verarbeitung der Daten

### 2.2.1 Vorverarbeiten & Zusammenfassen zu einem Gesamtdatensatz

**PSP-Code:** AP-221

**Verantwortlicher:** Niklas Koch

**Aufgabenstellung:** Die gesammelten und heruntergeladen Daten müssen zu einem Datensatz (in zwei Ausführungen) zusammengefasst werden, damit diese einfacher dem Rest des Projektteams zur Verfügung gestellt werden können. Außerdem vereinfacht dieser Schritt die Arbeit mit den Daten in späteren Arbeitspaketen.

**Zieldefinition:** Das Ziel ist die Erstellung eines einzigen Dataframes, der alle gesammelten Texte enthält, inklusive der Klasse, der sie zugehören. Danach soll aus diesem Dataframe ein Trainings- und ein Testdatensatz erstellt werden.

**Geplanter Zeitraum:** 24.05.23 - 07.06.23.

**Voraussetzung:** Es müssen Texte aus allen vier Klassen, die erkannt werden sollen, bereits gefunden und heruntergeladen sein.

**Hilfsmittel:** Die Programmiersprache Python und das Modul pandas.

**Risiken:** Sollten Fehler bei der Konkatination der verschiedenen Daten zu einem Dataframe auftreten, könnten die Texte nicht mehr mit der Zielvariable übereinstimmen, was später zu einem Modell führen, welches die Texte falsch klassifiziert.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Dieses Arbeitspaket baut direkt auf die vorherigen Arbeitspakete auf und bildet zeitgleich die Grundlage für alle folgenden Arbeitspakete.

**Abnahmebedingungen:** Es existiert ein Dataframe der alle Texte (inklusive ihrer Klasse) enthält. Dieser Dataframe wird aufgeteilt in einen Trainings- und einen Testdatensatz. Der Trainingsdatensatz enthält 2000 Einträge (500 pro Textklasse) und der Testdatensatz 800 Einträge (200 pro Textklasse).

### 2.2.2 Feature-Extraktion der Texte

**PSP-Code:** AP-222

**Verantwortlicher:** Niklas Koch und Jasmina Pascanovic

**Aufgabenstellung:** Aus den rohen Texten, die sich in den Dataframes befinden, sollen zusätzliche Features extrahiert werden. Diese können für tieferes Textverständnis und Visualisierungen genutzt werden, sowie für Signifikanztests, um eventuell zusätzliche Features für die Klassifikation zu gewinnen, oder die vermutliche Vorgehensweise eines potenziellen Modells in der Zukunft nachvollziehen zu können.

**Zieldefinition:** Aus den rohen Textinformationen werden folgende weitere Features als Extraspalten einem neuen Dataframe hinzugefügt:

- **sentences:** Eine Liste mit allen Sätzen des Textes (tokenisiert)
- **num\_sentences:** Anzahl der Sätze im Text
- **words\_with\_stopwords:** Eine Liste mit allen Wörtern (inklusive Füllwörter) des Textes (tokenisiert)
- **num\_words\_with\_stopwords:** Anzahl der Wörter im Text (inklusive Füllwörter)
- **words\_without\_stopwords:** Eine Liste mit allen Wörtern (exklusive Füllwörter) des Textes (tokenisiert)
- **num\_words\_without\_stopwords:** Anzahl der Wörter im Text (exklusive Füllwörter)
- **lemmas:** Die Wörter des Textes lemmatisiert, also in ihre Grundform umgewandelt.
- **stops:** Eine Liste mit allen gefunden Füllwörtern im Text

**Geplanter Zeitraum:** 07.06.23 - 14.06.23.

**Voraussetzung:** Die Daten müssen bereits in einem Dataframe zusammengeführt worden sein.

**Hilfsmittel:** Die Programmiersprache Python, das Modul pandas und das Modul spaCy.

**Risiken:** Es könnten Fehler beim Tokenisieren der Sätze oder Wörter auftreten, welche die oben genannten Features verfälschen würden.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Die Features werden in den Arbeitspaketen für Visualisierungen und Signifikanztests genutzt. Außerdem eventuell in allen Arbeitspaketen, die sich mit der Klassifikation beschäftigen.

**Abnahmebedingungen:** Für jeden Eintrag im Trainings- und Testdatensatz wurden die oben genannten Features extrahiert. Für die Arbeit mit diesen Features wurden zwei separate Dataframes erstellt.

### 2.2.3 Daten zur Verfügung stellen

**PSP-Code:** AP-223

**Verantwortlicher:** Niklas Koch

**Aufgabenstellung:** Alle erstellten Dataframes oder der Code, mit welchem die Daten erzeugt werden können, müssen dem Rest des Projektteams zur Verfügung gestellt werden. Dafür wird ein Github-Repository genutzt.

**Zieldefinition:** Das Nutzen von Github, um die Ergebnisse der vorherigen Arbeitspakete dem Rest des Teams zur Verfügung zustellen.

**Geplanter Zeitraum:** 14.06.23 - 16.06.23.

**Voraussetzung:** Die Arbeitspakete AP-211 bis AP-215 müssen abgearbeitet sein.

**Hilfsmittel:** Github.

**Risiken:** Daten können aus unerwarteten Gründen dem Rest des Teams nicht zur Verfügung gestellt werden.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Die Ergebnisse dieses Arbeitspakets sind die Grundlage des gesamten Projekts und basiert auf die Ergebnisse aller vorherigen Arbeitspakete.

**Abnahmebedingungen:** Die bearbeiteten Daten und der dafür genutzte Code wurden erfolgreich auf Github hochgeladen.

## 2.3 Visualisierungen

### 2.3.1 Grobe Datenanalyse (Data Exploration)

**PSP-Code:** AP-231

**Verantwortlicher:** Niklas Koch und Jasmina Pascanovic

**Aufgabenstellung:** Um ein gutes Verständnis über die Daten und unterschiedlichen Textklassen zu bekommen, wird eine grobe explorative Analyse der Daten durchgeführt.

**Zieldefinition:** Das Ziel ist es, bereits vor der Anwendung eines Machine Learning (ML) Algorithmus, Unterschiede zwischen oder Auffälligkeiten in den verschiedenen Textarten zu finden. Außerdem sollen Fehler, die beim Herunterladen, Konkatenieren oder der Feature-Extraction der Daten entstanden sein könnten, ausgeschlossen werden.

**Geplanter Zeitraum:** 07.06.23 - 20.06.23.

**Voraussetzung:** Die Vorverarbeitung der Daten, sowie das Extrahieren zusätzlicher Features.

**Hilfsmittel:** Die Programmiersprache Python, sowie die Module pandas und matplotlib.

**Risiken:** Fehler in den Daten oder besondere Auffälligkeiten fallen nicht auf.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Eine Schnittstelle besteht mit den vorherigen Arbeitspaketen und dem Arbeitspaket AP-232.

**Abnahmebedingungen:** Keine.

### 2.3.2 Visualisierung der zusätzlichen Features

**PSP-Code:** AP-232

**Verantwortlicher:** Jasmina Pascanovic

**Aufgabenstellung:** Mithilfe der in Arbeitspaket AP-222 erstellten zusätzlichen Features sollen Visualisierungen erstellt werden, die ein besseres Verständnis für die Daten und unterschiedlichen Texte ermöglichen.

**Zieldefinition:** Das Ziel ist es, mithilfe von Visualisierungen, ein besseres Verständnis für die verschiedenen Textarten zu bekommen.

**Geplanter Zeitraum:** 07.06.23 - 20.06.23.

**Voraussetzung:** Die Vorverarbeitung der Daten, sowie das Extrahieren zusätzlicher Features.

**Hilfsmittel:** Die Programmiersprache Python, sowie die Module pandas und matplotlib.

**Risiken:** Fehler in den Daten oder besondere Auffälligkeiten fallen nicht auf.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Eine Schnittstelle besteht mit den vorherigen Arbeitspaketen, sowie mit dem Arbeitspaket AP-31.

**Abnahmebedingungen:** Keine

## 3 Klassifikation

### 3.1 Signifikanztests für zusätzliche Features

**PSP-Code:** AP-31

**Verantwortlicher:** Jasmina Pascanovic

**Aufgabenstellung:** Es sollen Signifikanztests durchgeführt werden, welche den Einfluss der zusätzlichen Features auf die Zielvariable der Klassifikation beschreiben.

**Zieldefinition:** Die Signifikanztests sollen den Einfluss der extrahierten Features auf die Textklasse beschreiben.

**Geplanter Zeitraum:** 07.06.23 - 20.06.23.

**Voraussetzung:** Die zusätzlichen Features müssen extrahiert worden sein.

**Hilfsmittel:** Die Programmiersprache Python, das Modul pandas und das Modul matplotlib.

**Risiken:** Fehler in der Berechnung der Signifikanzen kann dazu führen, dass bei der Klassifikation Fehler gemacht werden.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Die Ergebnisse dieses Arbeitspakets spielen für die Arbeitspakete der Klassifikation eine Rolle. Sie basieren außerdem auf das vorherige Arbeitspaket der Feature Extraction.

**Abnahmebedingungen:** Signifikanztests wurden durchgeführt.

## 3.2 Umsetzung der Klassifikation mit simplem Algorithmus

**PSP-Code:** AP-32

**Verantwortlicher:** Niklas Koch

**Aufgabenstellung:** Die Klassifikation soll testweise mit einem simplen Algorithmus, wie einer Support-Vector-Machine durchgeführt werden.

**Zieldefinition:** Um die Vorgehensweise der Klassifikation besser zu verstehen und vorerst simpel zu halten, wird sie testweise mit einer simplen Methode, wie einer Support-Vector-Machine, durchgeführt.

**Geplanter Zeitraum:** 10.06.23 - 20.06.23.

**Voraussetzung:** Die Daten müssen vorhanden und zusammengefasst worden sein.

**Hilfsmittel:** Die Programmiersprache Python, die Module pandas und scikit-learn.

**Risiken:** Keine besonderen Risiken.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Die Ergebnisse werden eventuell im späteren Arbeitspaket beim Vergleich der Modelle genutzt.

**Abnahmebedingungen:** Klassifikation mit simplem Algorithmus wurde durchgeführt.



### 3.3 Umsetzung der Klassifikation mit eigenem neuronalen Netz

**PSP-Code:** AP-33

**Verantwortlicher:** Niklas Koch

**Aufgabenstellung:** Die Klassifikation soll mit einem neuronalen Netz durchgeführt werden.

**Zieldefinition** Ziel ist das Erstellen eines eigenen Neuronalen Netz zur Klassifikation der unterschiedlichen Textarten.

**Geplanter Zeitraum:** 07.06.23 - 20.06.23.

**Voraussetzung:** Die Daten müssen vorhanden und zusammengefasst worden sein.

**Hilfsmittel:** Die Programmiersprache Python, die Module pandas und tensorflow bzw. keras.

**Risiken:** Fehler beim Aufbau des Neuronalen Netz können zu Falschaussagen führen.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Die Ergebnisse werden eventuell im späteren Arbeitspaket beim Vergleich der Modelle genutzt.

**Abnahmebedingungen:** Klassifikation mit eigenem neuronalem Netz wurde durchgeführt.

### 3.4 Optimierung des eigenen neuronalen Netzes zur Klassifikation

**PSP-Code:** AP-34

**Verantwortlicher:** Niklas Koch

**Aufgabenstellung:** Das eigene Neuronale Netz soll optimiert werden, um Texte genauer und präziser einer bestimmten Klasse zuzuordnen oder auch „Unsicherheit“ auszudrücken.

**Zieldefinition:** Das eigene neuronale Netz soll genauere und präzisere Antworten ausgeben und besser funktionieren.

**Geplanter Zeitraum:** 20.06.23 - 30.06.23.

**Voraussetzung:** Die Daten müssen vorhanden und zusammengefasst worden sein. Das Neuronale Netz muss bereits aufgebaut worden sein.

**Hilfsmittel:** Die Programmiersprache Python, die Module pandas und tensorflow bzw. keras.

**Risiken:** Fehler beim Aufbau des Neuronalen Netz können zu Falschaussagen führen.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Die Ergebnisse werden eventuell im späteren Arbeitspaket beim Vergleich der Modelle genutzt.

**Abnahmebedingungen:** Optimierung des eigenen neuronalen Netzes wurde durchgeführt.

### 3.5 Umsetzung der Klassifikation mit einem Transformer-Modell

**PSP-Code:** AP-35

**Verantwortlicher:** Niklas Koch

**Aufgabenstellung:** Die Klassifikation soll mit einem bereits auf Texte spezifiziertem Transformer-Modell durchgeführt werden.

**Zieldefinition:** Ziel ist das Erstellen eines eigenen Transformer-Modells zur Klassifikation der unterschiedlichen Textarten.

**Geplanter Zeitraum:** 20.06.23 - 30.06.23.

**Voraussetzung:** Die Daten müssen vorhanden und zusammengefasst worden sein.

**Hilfsmittel:** Die Programmiersprache Python, die Module pandas und tensorflow bzw. keras, Hugging-Face.

**Risiken:** Komplikationen beim Nutzen eines Modells von Hugging-Face.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Die Ergebnisse werden eventuell im späteren Arbeitspaket beim Vergleich der Modelle genutzt.

**Abnahmebedingungen:** Klassifikation mit Transformer-Modell wurde durchgeführt.

### 3.6 Optimierung des Transformer-Modells zur Klassifikation

**PSP-Code:** AP-36

**Verantwortlicher:** Niklas Koch

**Aufgabenstellung:** Das Transformer-Modell soll optimiert werden, um Texte genauer und präziser einer bestimmten Klasse zuzuordnen oder auch „Unsicherheit“ auszudrücken.

**Zieldefinition:** Das Transformer-Modell soll genauere und präzisere Antworten ausgeben und besser funktionieren.

**Geplanter Zeitraum:** 30.06.23 - 06.07.23.

**Voraussetzung:** Die Daten müssen vorhanden und zusammengefasst worden sein. Das Neuronale Netz muss bereits aufgebaut worden sein.

**Hilfsmittel:** Die Programmiersprache Python, die Module pandas und tensorflow bzw. keras, Hugging-Face.

**Risiken:** Fehler beim Aufbau des Transformer-Modells können zu Falschaussagen führen.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Die Ergebnisse werden eventuell im späteren Arbeitspaket beim Vergleich der Modelle genutzt.

**Abnahmebedingungen:** Optimierung des Transformer-Modells wurde durchgeführt.

### 3.7 Vergleich der verschiedenen Verfahren auf Effektivität

**PSP-Code:** AP-37

**Verantwortlicher:** Niklas Koch

**Aufgabenstellung:** Die verschiedenen Modelle aus den vorherigen Arbeitspaketen sollen miteinander verglichen werden.

**Zieldefinition:** Durch den Vergleich der Modelle.

**Geplanter Zeitraum:** 06.07.23 - 08.07.23.

**Voraussetzung:** Die verschiedenen Modelle müssen aufgebaut werden.

**Hilfsmittel:** Die Programmiersprache Python, die Module pandas und tensorflow bzw. keras, Hugging-Face.

**Risiken:** Fehler beim Vergleich könnten zu Falschaussagen müssen.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Das beste bzw. am besten angepasste Modell aus dem Vergleich wird für das Endprodukt genutzt und in die UI integriert werden.

**Abnahmebedingungen:** Der Vergleich der Modelle wurde durchgeführt.

## 4 Zusammenfassung

### 4.1 Gestaltung einer Pipeline

#### 4.1.1 Recherche einer Trainingspipeline

**PSP-Code:** AP-411

**Verantwortlicher:** Niclas Cramer, Antoine Fuchs

**Aufgabenstellung:** Die primäre Aufgabenstellung besteht darin, verschiedene Methoden und Technologien für die Entwicklung von Trainingspipelines zu erforschen und zu analysieren. Die Recherche beinhaltet auch das Studium von Best Practices und aktuellen Trends in diesem Bereich.

**Zieldefinition:** Das Ziel dieses Arbeitspakets besteht darin, ein detailliertes Verständnis der verfügbaren Trainingspipelines zu erlangen. Dieses Wissen soll dazu dienen, eine fundierte Entscheidung darüber treffen zu können, welche Technologie für unser spezifisches Projekt am besten geeignet ist.

**Geplanter Zeitraum:** 24.05.23 - 31.05.23

**Voraussetzung:** Eine grundlegende Voraussetzung für die effektive Durchführung der Recherche ist ein Basisverständnis von maschinellem Lernen und Trainingspipelines.

**Hilfsmittel:** Zu den benötigten Hilfsmitteln gehören der Zugang zu technischen Dokumentationen, Online-Kursen und -Seminaren sowie entsprechenden Software- und Hardware-Ressourcen.

**Risiken:** Es gibt verschiedene Risiken, die berücksichtigt werden müssen. Dazu zählen der Mangel an verfügbaren Ressourcen für die Durchführung der Recherche, technologische Einschränkungen, die das Verständnis von bestimmten Trainingspipelines erschweren könnten, und potenzieller Zeitdruck.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Dieses Arbeitspaket stellt eine Schnittstelle zu anderen Arbeitspaketen darstellen, insbesondere in Bezug auf die Implementierung und Optimierung der letztlich ausgewählten Trainingspipeline.

**Abnahmebedingungen:** Die Abnahmebedingungen sind erfüllt, wenn die Ergebnisse der Recherche in einem umfassenden Bericht dokumentiert werden. Prüfung und Abnahme des Berichts erfolgt durch den Projektleiter.

### 4.1.2 Implementierung der Vorverarbeitung

**PSP-Code:** AP-412

**Verantwortlicher:** Niclas Cramer, Antoine Fuchs

**Aufgabenstellung:** Die Aufgabenstellung für dieses Arbeitspaket beinhaltet die Implementierung und Überprüfung der Daten-Vorverarbeitungsstrategien und -techniken für die Durchführung des Trainierens von Transformer-Modellen.

**Zieldefinition:** Das Ziel ist die erfolgreiche Implementierung von robusten und effizienten Daten-Vorverarbeitung, die dazu dienen, die Datenqualität und die Genauigkeit des maschinellen Lernens zu optimieren und die Nutzung der verschiedenen Modelle zu ermöglichen.

**Geplanter Zeitraum:** 31.05.23 - 07.06.23

**Voraussetzung:** Eine wesentliche Voraussetzung für dieses Arbeitspaket ist ein gründliches Verständnis von Daten-Vorverarbeitungstechniken und -strategien, sowie eine gewisse Erfahrung in der Anwendung solcher Methoden.

**Hilfsmittel:** Die erforderlichen Hilfsmittel umfassen Software-Werkzeuge zur Datenanalyse und -verarbeitung sowie Hardware-Ressourcen, um die Vorverarbeitungsprotokolle ausführen und testen zu können.

**Risiken:** Zu den potenziellen Risiken gehören technische Schwierigkeiten, die sich auf die Implementierung der Vorverarbeitungsprotokolle auswirken könnten, sowie unvorhergesehene Probleme mit den Daten, die Anpassungen der geplanten Vorverarbeitungsstrategien erfordern könnten.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Die Arbeit an diesem Paket wird voraussichtlich eng mit anderen Arbeitspaketen verbunden sein, insbesondere denen, die sich auf die spätere Verarbeitung und Analyse der vorverarbeiteten Daten beziehen.

**Abnahmebedingungen:** Die erfolgreiche Implementierung der Vorverarbeitungsprotokolle muss durch eine systematische Validierung und Überprüfung der Datenqualität nach der Vorverarbeitung nachgewiesen werden, in dem die Daten nach Verarbeitung überprüft werden.



### 4.1.3 Erstellen von Referenz-Zusammenfassungen für das Training der Modelle

**PSP-Code:** AP-413

**Verantwortlicher:** Niclas Cramer, Antoine Fuchs

**Aufgabenstellung:** Die Hauptaufgabe besteht darin, eine ausreichende Anzahl von qualitativ hochwertigen Referenz-Zusammenfassungen zu erstellen, die für das Training der maschinellen Lernmodelle im Rahmen von Supervised Learning verwendet werden können.

**Zieldefinition:** Das Ziel ist die Erstellung einer umfassenden Sammlung von Referenz-Zusammenfassungen, die die Vielfalt und Komplexität der Daten widerspiegeln.

**Geplanter Zeitraum:** 07.06.23 - 21.06.23

**Voraussetzung:** Eine wesentliche Voraussetzung für dieses Arbeitspaket ist ein tiefgreifendes Verständnis der Anforderungen an die Modelle sowie der Fähigkeiten und Einschränkungen des maschinellen Lernens.

**Hilfsmittel:** Die benötigten Hilfsmittel umfassen Software-Werkzeuge zur Datenverarbeitung und Texterstellung sowie Zugang zu den Daten, aus denen die Zusammenfassungen erstellt werden sollen. Diese können gleichermaßen mit bereits trainierten Modellen erstellt werden.

**Risiken:** Zu den möglichen Risiken gehören die Gefahr von Verzögerungen bei der Datensammlung und -verarbeitung und die Schwierigkeit, qualitativ hochwertige Zusammenfassungen zu erstellen, die den Anforderungen der Modelle gerecht werden. Dies kann schlussendlich zur Reduzierung der Qualität des Endproduktes führen

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Dieses Arbeitspaket bildet eine Schnittstelle zu den Arbeitspaketen bilden, die sich auf das Training und die Validierung der Modelle beziehen.

**Abnahmebedingungen:** Die erstellten Referenz-Zusammenfassungen müssen die festgelegten Qualitätsstandards erfüllen (Kompressionsrate und Inhalte des Ausgangstextes) und für das Training der Modelle geeignet sein. Die Abnahme erfolgt durch den Projektleiter, der die Qualität und Eignung der Zusammenfassungen überprüft. Hierbei müssen die Referenzzusammenfassungen für verschiedene Kompressionsraten erstellt werden.

#### 4.1.4 Erstellen der Trainingspipeline für die Modellerstellung

**PSP-Code:** AP-414

**Verantwortlicher:** Niclas Cramer, Antoine Fuchs

**Aufgabenstellung:** Die Hauptaufgabe besteht darin, eine effiziente und effektive Trainingspipeline für die Modellerstellung zu konzipieren und zu implementieren.

**Zieldefinition:** Das Ziel dieses Arbeitspakets ist die Bereitstellung einer funktionsfähigen und effizienten Trainingspipeline, die es erlaubt, maschinelle Lernmodelle systematisch zu trainieren und zu verbessern.

**Geplanter Zeitraum:** 07.06.23 - 14.06.23

**Voraussetzung:** Eine wesentliche Voraussetzung ist ein tiefgreifendes Verständnis der Prinzipien und Techniken des maschinellen Lernens sowie der Prozesse und Werkzeuge, die für die Erstellung von Trainingspipelines erforderlich sind.

**Hilfsmittel:** Die benötigten Hilfsmittel umfassen verschiedene Software-Werkzeuge zur Modellierung und zum maschinellen Lernen, Hardware-Ressourcen für das Training der Modelle, sowie Daten für das Training und die Validierung der Modelle.

**Risiken:** Potenzielle Risiken umfassen technische Probleme, die die Implementierung der Pipeline behindern könnten, sowie Schwierigkeiten bei der Anpassung der Pipeline an die spezifischen Anforderungen der Modelle.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Dieses Arbeitspaket wird voraussichtlich eine enge Verbindung zu den Arbeitspaketen haben, die sich auf das Training und die Verbesserung der Modelle, die Datenverarbeitung und die Modellvalidierung beziehen.

**Abnahmebedingungen:** Die erfolgreiche Implementierung der Trainingspipeline muss durch eine systematische Überprüfung und Validierung der Funktionsweise und Effizienz der Pipeline nachgewiesen werden. Die Abnahme erfolgt durch den Projektleiter, der die Pipeline bewertet und den Abschlussbericht prüft.

#### 4.1.5 Erstellen der Evaluationspipeline für die Kompressionsrate und den ROUGE-Score

**PSP-Code:** AP-415

**Verantwortlicher:** Niclas Cramer, Antoine Fuchs

**Aufgabenstellung:** Die Hauptaufgabe besteht darin, eine Evaluationspipeline zu erstellen, die in der Lage ist, sowohl die Kompressionsrate als auch den ROUGE-Score der erstellten Modelle zu bewerten. Diese Evaluationsmetriken helfen dabei, die Qualität und Leistung der Modelle zu messen und zu verbessern.

**Zieldefinition:** Das Ziel ist die Implementierung einer robusten und genauen Evaluationspipeline, die dazu dient, die Performance der maschinellen Lernmodelle systematisch zu bewerten und zu verbessern.

**Geplanter Zeitraum:** 14.06.23 - 25.06.23

**Voraussetzung:** Eine wesentliche Voraussetzung für dieses Arbeitspaket ist ein tiefes Verständnis von Evaluationsmetriken und -techniken im Bereich des maschinellen Lernens, insbesondere in Bezug auf die Kompressionsrate und den ROUGE-Score.

**Hilfsmittel:** Zu den benötigten Hilfsmitteln gehören Software-Tools zur Datenanalyse und -verarbeitung, zur Durchführung von Evaluationsprozessen und zur Berechnung der Evaluationsmetriken.

**Risiken:** Mögliche Risiken sind technische Schwierigkeiten bei der Implementierung der Evaluationspipeline, Probleme mit der Genauigkeit der Evaluationsmetriken sowie zeitliche Diskrepanzen.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Dieses Arbeitspaket besitzt Schnittstellen zu den Arbeitspaketen, die sich auf das Training der Modelle und die Modellverbesserung beziehen.

**Abnahmebedingungen:** Die Abnahme erfolgt nach einer systematischen Überprüfung der Implementierung der Evaluationspipeline und ihrer Fähigkeit, genaue und zuverlässige Bewertungen zu liefern. Ein Abschlussbericht, der die Implementierung der Pipeline, mögliche

Schwierigkeiten und die erzielten Ergebnisse dokumentiert, wird vom Projektleiter geprüft und abgenommen.

## 4.2 Gestaltung von eigens trainierten Modellen

### 4.2.1 Recherche von Transformer-Modellen

**PSP-Code:** AP-421

**Verantwortlicher:** Niclas Cramer, Antoine Fuchs

**Aufgabenstellung:** Die Aufgabenstellung für dieses Arbeitspaket besteht darin, eine umfassende Recherche zu aktuellen Transformer-Modellen im Bereich des maschinellen Lernens durchzuführen. Dies umfasst die Identifizierung relevanter Modelle, die Analyse ihrer Eigenschaften und Leistungen und die Bewertung ihrer Eignung für die geplanten Projektaufgaben.

**Zieldefinition:** Das Ziel dieser Recherche ist es, ein umfassendes Verständnis der verfügbaren Transformer-Modelle zu erlangen, ihre Stärken und Schwächen zu identifizieren und Empfehlungen für die Modelle abzugeben, die am besten für die Projektziele geeignet sind.

**Geplanter Zeitraum:** 24.05.23 - 31.05.23

**Voraussetzung:** Die Voraussetzung für dieses Arbeitspaket ist ein gründliches Verständnis der Prinzipien und Techniken des maschinellen Lernens und insbesondere der Funktionsweise und Anwendung von Transformer-Modellen.

**Hilfsmittel:** Die benötigten Hilfsmittel umfassen Zugang zu relevanten wissenschaftlichen Publikationen und Datenbanken, Software-Werkzeuge zur Datenanalyse und eventuell notwendige Rechenressourcen für das Testen und Experimentieren mit ausgewählten Modellen.

**Risiken:** Mögliche Risiken sind, dass die Recherche möglicherweise nicht alle relevanten Modelle abdeckt oder dass die bewerteten Modelle nicht den Anforderungen des Projekts entsprechen.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Die Ergebnisse dieser Recherche werden wahrscheinlich Einfluss auf mehrere andere Arbeitspakete haben, insbesondere diejenigen, die sich auf die Auswahl, das Training und die Validierung der Modelle beziehen.

**Abnahmebedingungen:** Die Abnahme erfolgt nach einer sorgfältigen Überprüfung der Rechercheergebnisse und der daraus gezogenen Schlussfolgerungen und Empfehlungen.

### 4.2.2 Testen der Basis-Transformer-Modelle für die Auswahl eines passenden Transformer-Modells

**PSP-Code:** AP-422

**Verantwortlicher:** Niclas Cramer, Antoine Fuchs

**Aufgabenstellung:** Die Aufgabe dieses Arbeitspakets ist das systematische Testen und Bewerten der in der vorhergehenden Recherche identifizierten Transformer-Modelle. Basierend auf diesen Tests soll das geeignetste Modell für die spezifischen Anforderungen des Projekts ausgewählt werden.

**Zieldefinition:** Das Ziel ist es, das Transformer-Modell zu identifizieren, das die besten Ergebnisse hinsichtlich Leistung, Effizienz und Anpassungsfähigkeit an die spezifischen Anforderungen des Projekts zeigt.

**Geplanter Zeitraum:** 31.05.23 - 14.06.23

**Voraussetzung:** Eine wichtige Voraussetzung für dieses Arbeitspaket ist ein umfassendes Verständnis der Transformer-Modelle und der Techniken und Werkzeuge, die für das Testen und Bewerten dieser Modelle erforderlich sind.

**Hilfsmittel:** Die benötigten Hilfsmittel umfassen Software-Werkzeuge für das maschinelle Lernen und die Datenanalyse, Hardware-Ressourcen für das Testen der Modelle und Daten für die Testprozesse.

**Risiken:** Zu den Risiken gehören technische Schwierigkeiten bei den Testprozessen, ungenaue Testergebnisse und Schwierigkeiten bei der Anpassung der Modelle an die Anforderungen des Projekts.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Dieses Arbeitspaket hat voraussichtlich Schnittstellen zu den Arbeitspaketen, die sich auf das Modelltraining, -validierung und -implementierung beziehen.

**Abnahmebedingungen:** Die Abnahme erfolgt nach einer sorgfältigen Überprüfung der Testprozesse und -ergebnisse und der getroffenen Modellauswahl

### 4.2.3 Trainieren der Modelle auf Basis der Trainingsdaten und Referenz-Zusammenfassungen

**PSP-Code:** AP-423

**Verantwortlicher:** Niclas Cramer, Antoine Fuchs

**Aufgabenstellung:** Die Hauptaufgabe in diesem Arbeitspaket besteht darin, die ausgewählten Transformer-Modelle unter Verwendung der bereitgestellten Trainingsdaten und Referenz-Zusammenfassungen zu trainieren.

**Zieldefinition:** Das Ziel dieses Arbeitspakets ist es, gut trainierte Modelle zu erzeugen, die in der Lage sind, effektive und genaue Textzusammenfassungen zu generieren.

**Geplanter Zeitraum:** 25.06.23 - 03.07.23

**Voraussetzung:** Voraussetzung für dieses Arbeitspaket ist ein umfassendes Verständnis der Techniken und Werkzeuge für das Training von Transformer-Modellen, sowie ein guter Überblick über die Trainingsdaten und Referenz-Zusammenfassungen.

**Hilfsmittel:** Zu den benötigten Hilfsmitteln gehören Software-Werkzeuge für das maschinelle Lernen, speziell für das Training von Transformer-Modellen, und ausreichende Hardware-Ressourcen für das Durchführen des Trainings. Zudem sind die Trainingsdaten und Referenz-Zusammenfassungen als Eingabe für das Training erforderlich.

**Risiken:** Mögliche Risiken sind technische Schwierigkeiten während des Trainings, unzureichende Leistung der trainierten Modelle und zeitliche Diskrepanzen.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Dieses Arbeitspaket hat voraussichtlich direkte Schnittstellen zu den Arbeitspaketen, die sich auf die Datenvorbereitung, Modellbewertung und Modellvalidierung beziehen.

**Abnahmebedingungen:** Die Abnahme erfolgt nach einer sorgfältigen Überprüfung der Trainingsprozesse und der Leistung der trainierten Modelle.

#### 4.2.4 Evaluierung der verschiedenen Modelle auf Basis des ROUGE-Scores

**PSP-Code:** AP-424

**Verantwortlicher:** Niclas Cramer, Antoine Fuchs

**Aufgabenstellung:** Die Hauptaufgabe dieses Arbeitspakets besteht darin, die Leistung der verschiedenen trainierten Modelle zu evaluieren, indem der ROUGE-Score berechnet wird. Dieser Score gibt einen quantitativen Hinweis auf die Qualität der von den Modellen erzeugten Zusammenfassungen.

**Zieldefinition:** Das Ziel dieses Arbeitspakets ist die objektive Bewertung und Vergleich der Leistung der verschiedenen Modelle, um das Modell mit der höchsten Leistung zu identifizieren.

**Geplanter Zeitraum:** 03.07.23 - 08.07.23

**Voraussetzung:** Eine wesentliche Voraussetzung für dieses Arbeitspaket ist ein tiefgehendes Verständnis der Evaluationsmetriken und -techniken im Bereich des maschinellen Lernens, insbesondere des ROUGE-Scores.

**Hilfsmittel:** Die benötigten Hilfsmittel umfassen Software-Tools zur Datenanalyse und -verarbeitung, zur Durchführung von Evaluationsprozessen und zur Berechnung des ROUGE-Scores.

**Risiken:** Mögliche Risiken sind technische Schwierigkeiten bei der Implementierung der Evaluationsprozesse, Probleme mit der Genauigkeit der Evaluationsmetriken und zeitliche Diskrepanzen.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Dieses Arbeitspaket steht in direktem Zusammenhang mit den Arbeitspaketen, die sich auf das Training der Modelle und die Modellverbesserung beziehen.

**Abnahmebedingungen:** Die Abnahme erfolgt nach einer systematischen Überprüfung der Evaluationsprozesse und der errechneten ROUGE-Scores. Ein Abschlussbericht, der die Durchführung der Evaluationsprozesse, die erzielten ROUGE-Scores und die daraus gezogenen Schlussfolgerungen dokumentiert.



### 4.2.5 Evaluierung der verschiedenen Modelle auf Basis der Kompressionsrate

**PSP-Code:** AP-425

**Verantwortlicher:** Niclas Cramer, Antoine Fuchs

**Aufgabenstellung:** Die Aufgabe in diesem Arbeitspaket besteht darin, die verschiedenen trainierten Modelle auf Basis ihrer Kompressionsrate zu evaluieren. Die Kompressionsrate gibt an, wie effektiv ein Modell einen Text in eine kürzere, aber dennoch aussagekräftige Form komprimieren kann.

**Zieldefinition:** Das Ziel dieses Arbeitspakets ist es, die Effizienz der verschiedenen Modelle in Bezug auf ihre Fähigkeit zur Textkompression zu bewerten. Hierdurch soll das Modell identifiziert werden, das eine gute Balance zwischen Inhaltserhaltung und Textreduktion bietet.

**Geplanter Zeitraum:** 03.07.23 - 08.07.23

**Voraussetzung:** Eine wichtige Voraussetzung für dieses Arbeitspaket ist ein umfassendes Verständnis der Methoden zur Messung und Beurteilung der Textkompression in maschinellen Lernmodellen.

**Hilfsmittel:** Die benötigten Hilfsmittel umfassen Software-Werkzeuge zur Textanalyse und -verarbeitung sowie zur Durchführung der Evaluationsprozesse.

**Risiken:** Mögliche Risiken sind technische Schwierigkeiten bei den Evaluationsprozessen, ungenaue Messungen der Kompressionsrate und zeitliche Diskrepanzen.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Dieses Arbeitspaket hat voraussichtlich direkte Schnittstellen zu den Arbeitspaketen, die sich auf das Training der Modelle und die Modellverbesserung beziehen.

**Abnahmebedingungen:** Die Abnahme erfolgt nach einer sorgfältigen Überprüfung der Evaluationsprozesse und der ermittelten Kompressionsraten. Ein Abschlussbericht, der die Durchführung der Evaluationsprozesse, die berechneten Kompressionsraten und die daraus gezogenen Schlussfolgerungen dokumentiert.

## 5 Entwicklung von zusätzlichen Modulen

### 5.1 Entwicklung einer Spracheingabemöglichkeit

#### 5.1.1 Recherche von Spracheingabemöglichkeiten

**PSP-Code:** AP-511

**Verantwortlicher:** Niclas Cramer, Antoine Fuchs

**Aufgabenstellung:** Die Aufgabe dieses Arbeitspakets besteht darin, verschiedene Technologien und Methoden zur Spracheingabe zu recherchieren und zu analysieren, um zu ermitteln, welche am besten für das Projekt geeignet sind.

**Zieldefinition:** Das Ziel dieses Arbeitspakets ist es, ein umfassendes Verständnis der verschiedenen Spracheingabemöglichkeiten zu erlangen und die am besten geeignete Methode für das Projekt zu identifizieren.

**Geplanter Zeitraum:** 05.06.23-12.06.23

**Voraussetzung:** Eine Voraussetzung für dieses Arbeitspaket ist ein grundlegendes Verständnis der Spracherkennungstechnologien und der Techniken zur Verarbeitung und Analyse von Spracheingaben.

**Hilfsmittel:** Die benötigten Hilfsmittel für dieses Arbeitspaket umfassen diverse Forschungs- und Analysewerkzeuge, Zugang zu wissenschaftlichen Publikationen und Technologie-Datenbanken, sowie möglicherweise Testhardware und Software für die praktische Evaluierung von Spracheingabetechnologien.

**Risiken:** Mögliche Risiken sind ungenaue oder unvollständige Rechercheergebnisse und Schwierigkeiten bei der Beurteilung der Eignung verschiedener Spracheingabemethoden für das Projekt.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Dieses Arbeitspaket hat Schnittstellen zu den Arbeitspaketen, die sich auf die Implementierung der Spracheingabefunktionalität und die Integration der Spracheingabe in das gesamte System beziehen.

**Abnahmebedingungen:** Die Abnahme erfolgt nach einer gründlichen Überprüfung der Rechercheergebnisse und der daraus gezogenen Schlussfolgerungen, welche vom Projektleiter geprüft und abgenommen wird.

### 5.1.2 Identifizieren von geeigneten Kommandos für die Spracheingabenutzung der Anwendung

**PSP-Code:** AP-512

**Verantwortlicher:** Niclas Cramer, Antoine Fuchs

**Aufgabenstellung:** Die Hauptaufgabe dieses Arbeitspakets besteht darin, eine Reihe von geeigneten und intuitiven Sprachkommandos für die Anwendung zu identifizieren und zu definieren.

**Zieldefinition:** Das Ziel dieses Arbeitspakets ist es, eine klare und einfach zu bedienende Spracheingabe zu erstellen, die den Nutzern der Anwendung hilft, die gewünschten Aktionen durchzuführen.

**Geplanter Zeitraum:** 19.06.23-26.06.23

**Voraussetzung:** Die Voraussetzung für dieses Arbeitspaket ist ein tiefes Verständnis der Spracheingabetechnologie und der spezifischen Anforderungen und Funktionalitäten der Anwendung.

**Hilfsmittel:** Die benötigten Hilfsmittel umfassen Werkzeuge für das Brainstorming und die Prototyping, Software für die Dokumentation und möglicherweise auch Nutzertests, um Feedback zur Benutzerfreundlichkeit und Intuitivität der definierten Kommandos zu erhalten.

**Risiken:** Die Risiken beinhalten die Möglichkeit, dass die identifizierten Kommandos nicht intuitiv oder leicht verständlich für die Nutzer sind, oder dass sie nicht gut mit der Funktionalität der Anwendung interagieren.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Dieses Arbeitspaket hat Schnittstellen zu den Arbeitspaketen, die sich auf die Implementierung der Spracheingabe und die Integration der Spracheingabe in die Anwendung beziehen.

**Abnahmebedingungen:** Die Abnahme erfolgt nach einer gründlichen Überprüfung der identifizierten Kommandos und ihrer Übereinstimmung mit den Anforderungen der Anwendung. Ein Abschlussbericht, der den Prozess der Identifizierung der Kommandos und deren Definition dokumentiert, wird vom Projektleiter geprüft und abgenommen.

### 5.1.3 Implementierung einer englischen Spracheingabefunktion

**PSP-Code:** AP-513

**Verantwortlicher:** Niclas Cramer, Antoine Fuchs

**Aufgabenstellung:** Die Hauptaufgabe dieses Arbeitspakets besteht darin, eine englische Spracheingabefunktion in die Anwendung zu implementieren, die es Nutzern ermöglicht, Befehle und Anfragen in englischer Sprache einzugeben.

**Zieldefinition:** Das Ziel dieses Arbeitspakets ist es, eine effiziente und benutzerfreundliche englische Spracheingabefunktion zu schaffen, die eine natürliche und intuitive Interaktion mit der Anwendung ermöglicht.

**Geplanter Zeitraum:** 19.06.23-26.06.23

**Voraussetzung:** Eine wesentliche Voraussetzung für dieses Arbeitspaket ist eine starke Kompetenz in der Programmierung und ein umfassendes Verständnis der Spracherkennungstechnologien, insbesondere im Hinblick auf die englische Sprache.

**Hilfsmittel:** Die benötigten Hilfsmittel umfassen Programmierwerkzeuge, Spracherkennungssoftware und eventuell Hardware für die Durchführung von Tests und die Verbesserung der Spracheingabefunktion.

**Risiken:** Mögliche Risiken sind technische Schwierigkeiten bei der Implementierung der Spracheingabe, Probleme mit der Erkennungsgenauigkeit oder der Benutzerfreundlichkeit und zeitliche Diskrepanzen.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Dieses Arbeitspaket hat Schnittstellen zu den Arbeitspaketen, die sich auf die Identifizierung geeigneter Kommandos und die Integration der Spracheingabe in die Anwendung beziehen.

**Abnahmebedingungen:** Die Abnahme erfolgt nach einer gründlichen Überprüfung und Testen der implementierten Spracheingabefunktion. Ein Abschlussbericht, der den Implementierungsprozess, die durchgeführten Tests und die erzielten Ergebnisse dokumentiert, wird vom Projektleiter geprüft und abgenommen.

#### 5.1.4 Testen der englischen Spracheingabefunktion

**PSP-Code:** AP-514

**Verantwortlicher:** Niclas Cramer, Antoine Fuchs

**Aufgabenstellung:** Die Hauptaufgabe dieses Arbeitspakets besteht darin, die neu implementierte englische Spracheingabefunktion zu testen, um ihre Leistung und Benutzerfreundlichkeit sicherzustellen.

**Zieldefinition:** Das Ziel dieses Arbeitspakets ist es, sicherzustellen, dass die englische Spracheingabefunktion korrekt funktioniert, die richtigen Befehle erkennt und die erwarteten Ergebnisse liefert.

**Geplanter Zeitraum:** 27.06.23-04.07.23

**Voraussetzung:** Eine Voraussetzung für dieses Arbeitspaket ist die erfolgreiche Implementierung der englischen Spracheingabefunktion in der vorhergehenden Phase.

**Hilfsmittel:** Die benötigten Hilfsmittel umfassen Testwerkzeuge und -software, sowie möglicherweise echte Benutzer oder künstlich generierte Daten für Usability-Tests.

**Risiken:** Mögliche Risiken beinhalten das Aufdecken von Fehlern oder Leistungsproblemen in der Spracheingabefunktion, die zu Verzögerungen bei der Inbetriebnahme führen können.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Dieses Arbeitspaket hat Schnittstellen zu den Arbeitspaketen, die sich auf die Implementierung und Verbesserung der englischen Spracheingabefunktion beziehen.

**Abnahmebedingungen:** Die Abnahme erfolgt nach einer gründlichen Überprüfung und Bestätigung, dass die englische Spracheingabefunktion korrekt funktioniert und die gestellten Anforderungen erfüllt. Ein Abschlussbericht, der die Testverfahren, die entdeckten Probleme und die vorgenommenen Korrekturen dokumentiert, wird vom Projektleiter geprüft und abgenommen.

## 5.2 Entwicklung einer Möglichkeit Dokumente einzulesen

### 5.2.1 Recherche von Möglichkeiten zum Einlesen von Dokumenten

**PSP-Code:** AP-521

**Verantwortliche:** Niclas Cramer, Antoine Fuchs

**Aufgabenstellung:** Durchführung einer umfassenden Recherche über verschiedene Möglichkeiten und Technologien zum Einlesen von Dokumenten, speziell für die Typen PDF, DOCX und TXT. Dies umfasst das Identifizieren von Tools, APIs oder Bibliotheken, die zur Extraktion und Verarbeitung von Texten aus diesen spezifischen Dokumentenformaten geeignet sind.

**Zieldefinition:** Ziel des Arbeitspakets ist es, spezifische Lösungen anzubieten, um Dokumente der Typen PDF, DOCX und TXT effizient einzulesen und den extrahierten Text für weitere Analysen nutzbar zu machen.

**Geplanter Zeitraum:** 05.06.23-12.06.23

**Voraussetzung:** Beim Umgang mit eingelesenem Text ist es wichtig, die Anforderungen und Ausgabeformate zu verstehen. Zudem benötigen wir grundlegende Kenntnisse über interne Strukturen und Merkmale von PDF, DOCX und TXT. Dies ermöglicht eine effektive Verarbeitung und Umwandlung des Textes.

**Hilfsmittel:** Keine.

**Risiken:** Risiken bei der Verarbeitung von Dokumenten umfassen die begrenzte Verfügbarkeit geeigneter Tools, Integrationsschwierigkeiten in die Arbeitsumgebung und Herausforderungen bei der Extraktion von Texten aus komplex strukturierten Dokumenten.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Datenextraktion und -vorbereitung

**Abnahmebedingungen:** Die Abnahmebedingungen für die Einlesung von PDF-, DOCX- und TXT-Dokumenten umfassen Folgendes: Eine Liste potenzieller Lösungen und Technologien wurde erstellt und dokumentiert. Die Vor- und Nachteile der identifizierten Möglichkeiten für jedes spezifische Dokumentenformat wurden bewertet und in einer Übersichtstabelle festgehalten. Zudem wurden Empfehlungen für die Auswahl der am besten geeigneten Lösung(en) für PDF, DOCX und TXT formuliert.

### 5.2.2 Implementierung des Einlesens von Dokumenten

**PSP-Code:** AP-522

**Verantwortliche:** Niclas Cramer, Antoine Fuchs

**Aufgabenstellung:** Implementierung einer Funktion zum Einlesen von Dokumenten der Typen PDF, DOCX und TXT unter Verwendung der ausgewählten Lösungen aus dem Arbeitspaket AP-521. Die Funktion soll die extrahierten Texte aus den Dokumenten extrahieren und für weitere Analysen verfügbar machen.

**Zieldefinition:** Ziel des Arbeitspakets ist es, eine funktionierende Implementierung für das Einlesen von PDF-, DOCX- und TXT-Dokumenten bereitzustellen, um den extrahierten Text für weitere Schritte in der Datenverarbeitung nutzbar zu machen.

**Geplanter Zeitraum:** 19.06.23-26.06.23

**Voraussetzung:** Eine Voraussetzung für die weiteren Schritte ist die Auswahl einer geeigneten Lösung für das Einlesen von PDF-, DOCX- und TXT-Dokumenten aus dem Arbeitspaket AP-521. Zudem ist ein grundlegendes Verständnis der Dokumentenstruktur und der Extraktionsanforderungen für jedes einzelne Format erforderlich.

**Hilfsmittel:** Die in AP-521 identifizierten Möglichkeiten zur Umsetzung des Dokumenteneinlesens.

**Risiken:** Bei der Implementierung sind einige Risiken zu beachten. Die Komplexität erhöht sich aufgrund der Vielfalt der Dokumentenformate und möglicher Strukturvariationen. Es können Herausforderungen auftreten, wenn es um den Umgang mit spezifischen Formattierungen oder Sonderzeichen in den Dokumenten geht.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Recherche von Möglichkeiten zum Einlesen von Dokumenten (AP-521).

**Abnahmebedingungen:** Die Implementierung zum Einlesen von PDF-, DOCX- und TXT-Dokumenten wurde erfolgreich abgeschlossen und umfassend getestet. Die Funktion extrahiert den Text präzise aus den Dokumenten und stellt ihn in einer nutzbaren Form für weitere



Analysen zur Verfügung. Eine detaillierte Dokumentation der Implementierung ist vorhanden, die sowohl Anweisungen zur Verwendung als auch potenzielle Einschränkungen enthält.

### 5.2.3 Testen auf verschiedenen Betriebssystemen und mit verschiedenen Dateiformaten

**PSP-Code:** AP-523

**Verantwortliche:** Niclas Cramer, Antoine Fuchs

**Aufgabenstellung:** Durchführung von Tests des implementierten Einlesevorgangs für Dokumente auf verschiedenen Betriebssystemen (z. B. Windows, macOS) und mit einer Vielzahl von Dateiformaten (PDF, DOCX, TXT). Ziel ist es, sicherzustellen, dass die Implementierung auf verschiedenen Plattformen und mit unterschiedlichen Dokumententypen reibungslos funktioniert.

**Zieldefinition:** Ziel des Arbeitspakets ist es, die Zuverlässigkeit und Kompatibilität des implementierten Einlesevorgangs für Dokumente auf verschiedenen Betriebssystemen und mit verschiedenen Dateiformaten zu gewährleisten.

**Geplanter Zeitraum:** 27.06.23-04.07.23

**Voraussetzung:** Für den weiteren Fortschritt sind zwei Voraussetzungen erforderlich: eine erfolgreiche Implementierung des Einlesevorgangs gemäß Arbeitspaket AP-522 und die Verfügbarkeit von Testdaten in verschiedenen Dateiformaten (PDF, DOCX, TXT).

**Hilfsmittel:** Testdateien verschiedener Dateiformate sowie Hardware mit unterschiedlichen Betriebssystemen.

**Risiken:** Bei der Umsetzung gibt es einige Risiken zu beachten. Dazu gehören das unterschiedliche Verhalten des Einlesevorgangs auf verschiedenen Betriebssystemen, Probleme bei der Verarbeitung spezifischer Dateiformate und Herausforderungen bei der Fehlererkennung und -behebung auf verschiedenen Plattformen.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Implementierung des Einlesens von Dokumenten (AP-522).

**Abnahmebedingungen:** Die Abnahmebedingungen für den Einlesevorgang von Dokumenten umfassen Folgendes: Der Einlesevorgang muss zuverlässig und korrekt auf verschiedenen Betriebssystemen wie Windows, macOS und Linux funktionieren. Zudem sollten

keine signifikanten Unterschiede im Verhalten oder in der Performance beim Einlesen unterschiedlicher Dateiformate wie PDF, DOCX und TXT auftreten. Um mögliche Probleme oder Einschränkungen zu identifizieren, wurden umfassende Tests durchgeführt, die potenzielle Schwierigkeiten auf verschiedenen Plattformen und mit verschiedenen Dateiformaten aufdecken und dokumentieren sollen.

## 6 Frontend

### 6.1 Anbindungen

#### 6.1.1 Anbindung von trainierten Modellen an das Frontend

**PSP-Code:** AP-611

**Verantwortliche:** Niclas Cramer, Antoine Fuchs, Niklas Koch, Jasmina Pascanovic

**Aufgabenstellung:** Implementierung der notwendigen Funktionalität, um trainierte Modelle an das Frontend anzubinden. Dies beinhaltet die Integration der Modelle in das bestehende Frontend-Framework, die Bereitstellung von Schnittstellen für die Kommunikation zwischen Frontend und Modellen sowie die Implementierung der Logik zur Anforderung von Vorhersagen oder Analysen von Benutzerdaten.

**Zieldefinition:** Ziel des Arbeitspakets ist es, eine nahtlose Kommunikation zwischen dem Frontend und den trainierten Modellen zu ermöglichen, um Benutzerdaten effektiv zu verarbeiten und Vorhersagen oder Analysen zurückzugeben.

**Geplanter Zeitraum:** 09.07.23 - 16.07.23.

**Voraussetzung:** Die Voraussetzungen für den weiteren Fortschritt sind das Vorhandensein und das erfolgreiche Training der Modelle. Zudem ist es wichtig, die Anforderungen des Frontends und der Schnittstellen für die Datenkommunikation zu verstehen.

**Hilfsmittel:** Die benötigten Hilfsmittel umfassen verschiedene Software-Werkzeuge wie die einzelnen Bibliotheken der erstellten Modelle sowie die zur Erstellung des Frontend genutzten Frameworks.

**Risiken:** Bei der Integration von Modellen in das Frontend-Framework bestehen Risiken in Bezug auf die Komplexität der Integration, Schwierigkeiten bei der Schnittstellenimplementierung und potenzielle Leistungs- oder Skalierungsprobleme bei der Verarbeitung großer Datenmengen im Frontend. Zusätzliche Schnittstellen zu anderen Arbeitspaketen sollten ebenfalls berücksichtigt werden.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Modelltraining und Frontend-Entwicklung.

**Abnahmebedingungen:** Die Abnahmebedingungen für die Integration der Modelle in das Frontend sind wie folgt: Die trainierten Modelle müssen erfolgreich in das Frontend integriert werden und über entsprechende Schnittstellen aufrufbar sein. Das Frontend sollte in der Lage sein, Anfragen an die Modelle zu senden und die erhaltenen Vorhersagen oder Analysen auf geeignete Weise darzustellen. Die Integration muss gründlich getestet werden, um eine zuverlässige und effiziente Kommunikation zwischen dem Frontend und den Modellen sicherzustellen.

### 6.1.2 Anbindung einer Spracheingabe

**PSP-Code:** AP-612

**Verantwortlicher:** Niclas Cramer, Antoine Fuchs, Niklas Koch, Jasmina Pascanovic

**Aufgabenstellung:** Implementierung der Funktionalität zur Anbindung einer Spracheingabe an das System. Dies beinhaltet die Integration eines Spracherkennungsmoduls oder einer Spracherkennungs-API, um gesprochene Sprache in Text umzuwandeln und für weitere Verarbeitungsschritte verfügbar zu machen.

**Zieldefinition:** Ziel des Arbeitspakets ist es, die Möglichkeit zur Spracheingabe in das System zu integrieren, um Benutzeranfragen oder Befehle per Sprache entgegennehmen zu können.

**Geplanter Zeitraum:** 09.07.23 - 16.07.23.

**Voraussetzung:** Um den Fortschritt zu ermöglichen, müssen zwei Voraussetzungen erfüllt sein. Erstens ist die Auswahl einer geeigneten Spracherkennungstechnologie oder einer Spracherkennungs-API erforderlich, die den Anforderungen des Systems entspricht. Zweitens ist ein umfassendes Verständnis der Anforderungen an die Spracheingabe-Funktionalität im System erforderlich.

**Hilfsmittel:** Die benötigten Hilfsmittel umfassen verschiedene Software-Werkzeuge wie die einzelnen Bibliotheken der Spracheingabefunktion sowie die zur Erstellung des Frontend genutzten Frameworks.

**Risiken:** Die Risiken im Zusammenhang mit der Spracheingabe sind die Komplexität der Integration in das bestehende System, Schwierigkeiten bei der Umsetzung der Spracherkennung und Umwandlung von Sprache in Text sowie potenzielle Herausforderungen bei der Behandlung von Sprachvarianz oder Hintergrundgeräuschen.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Frontend-Entwicklung und Entwicklung der Spracheingabe

**Abnahmebedingungen:** Die Abnahmebedingungen für die Spracheingabe-Funktionalität umfassen Folgendes: Die Integration der Funktion in das System ist erfolgreich abgeschlossen, sodass eine Umwandlung von gesprochener Sprache in Text ermöglicht wird. Das System ist in der Lage, die Spracheingabe zu akzeptieren, zu verarbeiten und entsprechende Aktionen auszulösen. Die Integration wurde umfangreich getestet und zeigt eine zuverlässige und genaue Umwandlung von Sprache in Text.

### 6.1.3 Anbindung der Dokumente-Einlesen-Funktion

**PSP-Code:** AP-613

**Verantwortlicher:** Niclas Cramer, Antoine Fuchs, Niklas Koch, Jasmina Pascanovic

**Aufgabenstellung:** Integration der implementierten Funktion zum Einlesen von Dokumenten in das System. Dies beinhaltet die Anbindung der Funktion an das Frontend, um Benutzern das Hochladen von Dokumenten zu ermöglichen, sowie die Verarbeitung der hochgeladenen Dokumente und Extraktion des Textinhalts für weitere Verarbeitungsschritte.

**Zieldefinition:** Ziel des Arbeitspakets ist es, die Funktionalität zum Einlesen von Dokumenten nahtlos in das System zu integrieren und Benutzern die Möglichkeit zu bieten, Dokumente hochzuladen und den extrahierten Text für weitere Analysen zu verwenden.

**Geplanter Zeitraum:** 09.07.23 - 16.07.23.

**Voraussetzung:** Um den Fortschritt sicherzustellen, sind zwei Voraussetzungen zu beachten. Erstens sollte das Einlesen von Dokumenten gemäß Arbeitspaket AP-522 erfolgreich implementiert worden sein. Zweitens ist ein gründliches Verständnis der Anforderungen und Schnittstellen des Frontends für den Upload von Dokumenten erforderlich. Durch Erfüllen dieser Voraussetzungen kann der reibungslose Fortschritt gewährleistet werden.

**Risiken:** Die Risiken im Zusammenhang mit dem Einlesen von Dokumenten sind die Komplexität der Integration der Einlesefunktion in das Frontend und die Systemarchitektur, Herausforderungen bei der Handhabung verschiedener Dokumentenformate und -größen sowie potenzielle Auswirkungen auf die Systemleistung bei der Verarbeitung großer Dokumente. Es ist wichtig, diese Risiken zu berücksichtigen und entsprechende Maßnahmen zu ergreifen, um eine erfolgreiche Einbindung der Einlesefunktion zu gewährleisten und potenziellen Herausforderungen zu begegnen.

**Hilfsmittel:** Die benötigten Hilfsmittel umfassen verschiedene Software-Werkzeuge wie die einzelnen Bibliotheken der Dokumenteinlesefunktion sowie die zur Erstellung des Frontend genutzten Frameworks.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Frontend-Entwicklung und Modulentwicklung



**Abnahmebedingungen:** Die Abnahmebedingungen für die Einlesefunktionalität von Dokumenten sind wie folgt: Die Integration der Funktion in das Frontend wurde erfolgreich abgeschlossen und ermöglicht Benutzern das Hochladen von Dokumenten. Die hochgeladenen Dokumente werden erfolgreich verarbeitet, und der extrahierte Text steht für weitere Analysen oder Verarbeitungsschritte zur Verfügung. Die Integration wurde umfassend getestet und zeigt eine zuverlässige und effiziente Verarbeitung der hochgeladenen Dokumente. Durch das Erfüllen dieser Abnahmebedingungen wird gewährleistet, dass die Einlesefunktionalität den Anforderungen des Systems entspricht und fehlerfrei funktioniert.

## 6.2 Ausgabe

### 6.2.1 Ausgabe des Klassifikationsergebnisses im Frontend

**PSP-Code:** AP-621

**Verantwortlicher:** Niclas Cramer, Antoine Fuchs, Niklas Koch, Jasmina Pascanovic

**Aufgabenstellung:** Implementierung der Funktionalität zur Ausgabe des Klassifikationsergebnisses im Frontend. Dies beinhaltet die Integration der Klassifikationslogik in das Frontend, um Benutzern das Einreichen von Texten zur Klassifizierung zu ermöglichen und das Ergebnis der Klassifikation anzuzeigen.

**Zieldefinition:** Ziel des Arbeitspakets ist es, das Klassifikationsergebnis im Frontend anzuzeigen und den Benutzern eine klare Rückmeldung über die Klassifizierung ihrer Texte zu geben.

**Geplanter Zeitraum:** 09.07.23 - 16.07.23.

**Voraussetzung:** Um den Fortschritt sicherzustellen, müssen zwei Voraussetzungen erfüllt sein. Erstens sollte ein trainiertes Klassifikationsmodell vorhanden sein, das für die gewünschten Aufgaben geeignet ist. Zweitens ist ein umfassendes Verständnis der Anforderungen des Frontends für die Anzeige des Klassifikationsergebnisses erforderlich, einschließlich des Formats, der Darstellungsmöglichkeiten und der Integration in die Benutzeroberfläche.

**Hilfsmittel:** Die benötigten Hilfsmittel umfassen verschiedene Software-Werkzeuge wie die einzelnen Bibliotheken der erstellten Modelle sowie die zur Erstellung des Frontend genutzten Frameworks.

**Risiken:** Die Risiken im Zusammenhang mit der Klassifikationsfunktionalität bestehen in der Komplexität der Integration der Klassifikationslogik in das Frontend sowie in den Herausforderungen bei der Darstellung des Klassifikationsergebnisses in einer benutzerfreundlichen Form. Es ist wichtig, diese Risiken zu berücksichtigen und entsprechende Maßnahmen zu ergreifen, um eine erfolgreiche Integration der Klassifikationslogik zu gewährleisten und sicherzustellen, dass das Klassifikationsergebnis auf eine Weise präsentiert wird, die für die Benutzer leicht verständlich und benutzerfreundlich ist.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Frontend-Entwicklung und Erstellung der Klassifizierungsmodelle

**Abnahmebedingungen:** Die Abnahmebedingungen für die Klassifikationsfunktionalität sind wie folgt: Das Klassifikationsergebnis wird erfolgreich im Frontend angezeigt, basierend auf den vom Benutzer eingereichten Texten. Die Klassifikationslogik ist korrekt implementiert und liefert genaue und verlässliche Ergebnisse. Die Integration wurde umfassend getestet und zeigt eine zuverlässige Darstellung des Klassifikationsergebnisses im Frontend.

### 6.2.2 Ausgabe des Zusammenfassungsergebnisses im Frontend

**PSP-Code:** AP-622

**Verantwortlicher:** Niclas Cramer, Antoine Fuchs, Niklas Koch, Jasmina Pascanovic

**Aufgabenstellung:** Implementierung der Funktionalität zur Ausgabe des Zusammenfassungsergebnisses im Frontend. Dies beinhaltet die Integration der Zusammenfassungslogik in das Frontend, um Benutzern die Möglichkeit zu geben, Texte zur Zusammenfassung einzureichen und die generierte Zusammenfassung anzuzeigen.

**Zieldefinition:** Ziel des Arbeitspakets ist es, das Zusammenfassungsergebnis im Frontend anzuzeigen und den Benutzern eine prägnante Zusammenfassung ihrer Texte bereitzustellen.

**Geplanter Zeitraum:** 09.07.23 - 16.07.23

**Voraussetzung:** Um den Fortschritt sicherzustellen, müssen zwei Voraussetzungen erfüllt sein. Erstens sollte eine implementierte Zusammenfassungslogik oder ein trainiertes Zusammenfassungsmodell vorhanden sein, das für die gewünschten Aufgaben geeignet ist. Zweitens ist ein umfassendes Verständnis der Anforderungen des Frontends für die Anzeige des Zusammenfassungsergebnisses erforderlich, einschließlich des Formats, der Länge und der Integration in die Benutzeroberfläche.

**Hilfsmittel:** Die benötigten Hilfsmittel umfassen verschiedene Software-Werkzeuge wie die einzelnen Bibliotheken der erstellten Modelle sowie die zur Erstellung des Frontend genutzten Frameworks.

**Risiken:** Die Risiken im Zusammenhang mit der Zusammenfassungsfunktionalität bestehen in der Komplexität der Integration der Zusammenfassungslogik in das Frontend sowie in den Herausforderungen bei der Darstellung der Zusammenfassung in einer benutzerfreundlichen Form.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Frontend-Entwicklung und Erstellung des Zusammenfassungsmodell

**Abnahmebedingungen:** Das Zusammenfassungsergebnis wird erfolgreich im Frontend angezeigt, basierend auf den vom Benutzer eingereichten Texten. Die implementierte Zusammenfassungslogik ist korrekt und liefert prägnante und sinnvolle Zusammenfassungen. Die Integration wurde umfassend getestet und zeigt eine zuverlässige Darstellung des Zusammenfassungsergebnisses im Frontend.

### 6.2.3 Integration der Pipeline zur Durchführung der Klassifikation und Zusammenfassung im Frontend

**PSP-Code:** AP-623

**Verantwortlicher:** Niclas Cramer, Antoine Fuchs, Niklas Koch, Jasmina Pascanovic

**Aufgabenstellung:** Integration der Pipeline zur Durchführung der Klassifikation und Zusammenfassung in das System. Dies beinhaltet die Einrichtung einer geeigneten Umgebung, in welcher die Zusammenfassung und Klassifikation korrekt angesteuert werden. Dies beinhaltet die notwendigen Skripte oder Workflows zur Datenverarbeitung, Modelldurchführung und -evaluation.

**Zieldefinition:** Ziel des Arbeitspakets ist es, die Pipeline zur Durchführung der Zusammenfassung und Klassifikation erfolgreich in das System und Frontend zu integrieren.

**Geplanter Zeitraum:** 09.07.23 - 16.07.23

**Voraussetzung:** Erstens sollte eine Trainingspipeline existieren und erfolgreich implementiert worden sein, um als Maßstab für die Umsetzung zu dienen. Diese Pipeline sollte in der Lage sein, Daten zu verarbeiten, Modelle zu trainieren, zu aktualisieren und zu evaluieren. Zweitens ist ein umfassendes Verständnis der Anforderungen an die Datenverarbeitung, Modellaktualisierung und -evaluation erforderlich. Dies beinhaltet die Anforderungen an die Datenquelle, die erforderlichen Datenverarbeitungsschritte, den Zeitplan für die Modellaktualisierung sowie die Bewertungsmaßstäbe und -metriken für die Modellbewertung.

**Hilfsmittel:** Die benötigten Hilfsmittel umfassen verschiedene Software-Werkzeuge wie die einzelnen Bibliotheken der erstellten Modelle sowie die zur Erstellung des Frontend genutzten Frameworks.

**Risiken:** Die Risiken im Zusammenhang mit der Trainingspipeline bestehen in der Komplexität der Integration in das bestehende System, potenziellen Herausforderungen bei der Skalierung für große Datensätze oder komplexe Modelle sowie Schwierigkeiten bei der Synchronisierung der Modelle mit anderen Systemkomponenten.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Modelltraining und Evaluation

---

**Abnahmebedingungen:** Die Pipeline ist erfolgreich in das System und Frontend integriert und angebunden. Sie ermöglicht die automatische Durchführung der Klassifikation und Zusammenfassung. Die Datenverarbeitung, Modelldurchführung und -evaluation erfolgen korrekt und liefern qualitativ hochwertige und leistungsfähige Ergebnisse. Die Integration wurde umfassend getestet und zeigt eine zuverlässige und effiziente Durchführung.

## 6.3 Barrierefreiheit

### 6.3.1 Recherche für ein barrierefreies Frontend

**PSP-Code:** AP-631

**Verantwortlicher:** Niklas Koch, Jasmina Pascanovic

**Aufgabenstellung:** Durchführung einer umfassenden Recherche über die Anforderungen und bewährten Methoden für ein barrierefreies Frontend. Dies umfasst die Untersuchung von Richtlinien, Standards und Techniken zur Gewährleistung der Zugänglichkeit und Benutzbarkeit des Frontends für Menschen mit verschiedenen Arten von Behinderungen.

**Zieldefinition:** Ziel des Arbeitspakets ist es, das Wissen und das Verständnis über barrierefreies Design und Entwicklung im Frontend zu erweitern und Empfehlungen zur Umsetzung eines barrierefreien Frontends zu formulieren.

**Geplanter Zeitraum:** 27.06.23 - 04.07.23

**Voraussetzung** Keine

**Risiken:** Die Risiken im Zusammenhang mit der Umsetzung eines barrierefreien Frontends bestehen in der Komplexität und Vielfalt der Anforderungen, dem Verständnis der technischen Aspekte und Implementierungsdetails für barrierefreies Design sowie den Herausforderungen bei der Umsetzung der Empfehlungen in das bestehende Frontend-Framework oder Design.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Frontend-Entwicklung

**Abnahmebedingungen:** Die Abnahmebedingungen für ein barrierefreies Frontend sind, dass eine umfassende Recherche über die Anforderungen und bewährten Methoden für barrierefreies Design durchgeführt und dokumentiert wurde. Zudem wurden Empfehlungen und Richtlinien zur Umsetzung eines barrierefreien Frontends formuliert und stehen als Grundlage für die weitere Entwicklung zur Verfügung. Darüber hinaus wurde das erweiterte Wissen über barrierefreies Design und Entwicklung im Frontend dokumentiert und kann in der Entwicklung und Implementierung berücksichtigt werden.



### 6.3.2 Design eines barrierefreien Frontends

**PSP-Code:** AP-632

**Verantwortlicher:** Niklas Koch, Jasmina Pascanovic

**Aufgabenstellung:** Design und Entwicklung eines barrierefreien Frontends, das den Richtlinien und Empfehlungen für eine verbesserte Zugänglichkeit und Benutzbarkeit für Menschen mit verschiedenen Behinderungen entspricht. Dies beinhaltet die Gestaltung von benutzerfreundlichen Schnittstellen, die Verwendung geeigneter Farben, Kontraste und Schriftgrößen, die Implementierung von Tastaturnavigation und Unterstützung für Bildschirmleser sowie die Einbindung von Hilfstechnologien wie ARIA-Tags (Accessible Rich Internet Applications).

**Zieldefinition:** Ziel des Arbeitspakets ist es, ein barrierefreies Frontend zu entwerfen und zu entwickeln, das allen Benutzern eine gleichberechtigte und effektive Nutzung der Anwendung ermöglicht.

**Geplanter Zeitraum:** 04.07.23 - 11.07.23

**Voraussetzung:** Erstens sollten die Ergebnisse der Recherche über barrierefreies Design und Entwicklung aus dem Arbeitspaket AP-631 vorliegen und analysiert worden sein. Diese Ergebnisse dienen als Grundlage für das Verständnis der Anforderungen und bewährten Praktiken im Bereich der Barrierefreiheit. Zweitens ist ein umfassendes Verständnis der Benutzeranforderungen und Zielgruppen der Anwendung erforderlich. Dies umfasst das Wissen über die Bedürfnisse und Einschränkungen der Benutzer, insbesondere im Hinblick auf Barrierefreiheit.

**Risiken:** Die Risiken im Zusammenhang mit der Umsetzung eines barrierefreien Frontends bestehen in der Komplexität der Umsetzung der barrierefreien Designprinzipien, den Herausforderungen bei der Sicherstellung der Benutzbarkeit für verschiedene Behinderungen und Hilfstechnologien sowie den möglichen Auswirkungen auf das visuelle Design und die ästhetische Gestaltung der Benutzeroberfläche.

**Hilfsmittel:** Prototyping-Tools wie XD oder Figma.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Frontend-Entwicklung

**Abnahmebedingungen:** Das barrierefreie Frontend wurde erfolgreich entworfen und entwickelt, basierend auf den Richtlinien und Empfehlungen für barrierefreies Design. Es erfüllt die Anforderungen der Zugänglichkeit und Benutzbarkeit für Menschen mit verschiedenen Arten von Behinderungen. Das Frontend wurde umfassend getestet und zeigt eine gute Benutzererfahrung für alle Benutzergruppen, unabhängig von ihren Fähigkeiten oder Hilfstechnologien.

### 6.3.3 Umsetzung und Implementierung des barrierefreien Frontend-Designs

**PSP-Code:** AP-633

**Verantwortlicher:** Niklas Koch, Jasmina Pascanovic

**Aufgabenstellung:** Umsetzung und Implementierung des zuvor entworfenen barrierefreien Frontend-Designs. Dies beinhaltet die Entwicklung der erforderlichen HTML-, CSS- und JavaScript-Codekomponenten gemäß den Richtlinien und Empfehlungen für barrierefreies Design. Es sollen die erforderlichen Interaktionsmöglichkeiten, Farben, Kontraste, Schriftgrößen und Unterstützung für Hilfstechnologien wie Bildschirmleser oder Tastaturnavigation implementiert werden.

**Zieldefinition:** Ziel des Arbeitspakets ist es, das barrierefreie Frontend-Design erfolgreich umzusetzen und zu implementieren, um eine zugängliche und benutzerfreundliche Anwendung zu gewährleisten.

**Geplanter Zeitraum:** 04.07.23 - 16.07.23

**Voraussetzung:** Um mit der Umsetzung des barrierefreien Frontends fortzufahren, müssen zwei Voraussetzungen erfüllt sein. Erstens sollte das Design des barrierefreien Frontends aus dem Arbeitspaket AP-632 erfolgreich abgeschlossen worden sein. Dieses Design bildet die Grundlage für die visuelle Gestaltung und die Benutzerinteraktionen im Frontend unter Berücksichtigung der Anforderungen des barrierefreien Designs. Zweitens ist ein umfassendes Verständnis der Anforderungen des Frontends und der technischen Umsetzung erforderlich. Dies beinhaltet das Wissen über die Funktionen, Interaktionsmöglichkeiten, Datenverarbeitung und die Integration mit anderen Systemkomponenten.

**Hilfsmittel:** Streamlit als Frontend-Framework

**Risiken:** Die Risiken im Zusammenhang mit der Umsetzung des barrierefreien Frontends bestehen in der Komplexität der Umsetzung des barrierefreien Designs in Codekomponenten, den Herausforderungen bei der Gewährleistung der korrekten Funktionalität und Benutzbarkeit für verschiedene Behinderungen und Hilfstechnologien sowie den potenziellen Auswirkungen auf die Performance oder Kompatibilität des Frontends mit verschiedenen Browsern oder Plattformen.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Frontend-Entwicklung

---

**Abnahmebedingungen:** Das barrierefreie Frontend wurde erfolgreich umgesetzt und implementiert, basierend auf dem zuvor entworfenen Design und den Richtlinien für barrierefreies Design. Das Frontend zeigt eine gute Performance und Kompatibilität mit verschiedenen Browsern und Plattformen.

## 6.4 Testen der gesamten Frontend-Anwendung

**PSP-Code:** AP-64

**Verantwortlicher:** Niclas Cramer, Antoine Fuchs, Niklas Koch, Jasmina Pascanovic

**Aufgabenstellung:** Durchführung umfangreicher Tests der gesamten Frontend-Anwendung, um die Funktionalität, Benutzerfreundlichkeit und Leistung sicherzustellen. Dies beinhaltet das Testen der einzelnen Funktionen, das Überprüfen der Benutzerinteraktion, das Durchführen von Usability-Tests und das Identifizieren und Beheben von etwaigen Fehlern oder Problemen.

**Zieldefinition:** Ziel des Arbeitspakets ist es, sicherzustellen, dass die gesamte Frontend-Anwendung qualitativ hochwertig ist, eine gute Benutzererfahrung bietet und reibungslos funktioniert.

**Geplanter Zeitraum:** 16.07.23 - 21.07.23.

**Voraussetzung:** Um mit der Implementierung des barrierefreien Frontends fortzufahren, müssen zwei Voraussetzungen erfüllt sein. Erstens sollte das barrierefreie Frontend-Design gemäß Arbeitspaket AP-633 erfolgreich implementiert worden sein. Dies bedeutet, dass das visuelle Design und die Benutzerinteraktionen entsprechend den Anforderungen des barrierefreien Designs umgesetzt wurden. Zweitens ist ein umfassendes Verständnis der Anforderungen der Frontend-Funktionen und Benutzerinteraktion erforderlich. Dies beinhaltet das Wissen über die verschiedenen Funktionen, Interaktionsmöglichkeiten und Datenverarbeitung im Frontend, um sicherzustellen, dass das barrierefreie Frontend diese Anforderungen erfüllt.

**Hilfsmittel:** Keine.

**Risiken:** Die Risiken im Zusammenhang mit der Implementierung des barrierefreien Frontends bestehen in potenziellen Fehlern oder Problemen, Herausforderungen bei der Identifizierung und Reproduktion von Fehlern während der Tests sowie Performance- oder Kompatibilitätsproblemen mit verschiedenen Browsern oder Plattformen.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** - Frontend-Entwicklung - Umsetzung und Implementierung des barrierefreien Frontend-Designs (AP-633)

**Abnahmebedingungen:** Die Abnahmebedingungen für die Frontend-Anwendung sind: umfassende Tests zur Gewährleistung einwandfreier Funktionalität, Erfüllung der Anforderungen an Benutzerfreundlichkeit und Zugänglichkeit, Behebung von Fehlern für Stabilität, sowie gute Performance und Kompatibilität auf verschiedenen Browsern und Plattformen.

## 7 Dokumentation

### 7.1 Dokumentation der allgemeinen Aspekte

#### 7.1.1 Dokumentation der Datenbeschaffung

**PSP-Code:** AP-711

**Verantwortlicher:** Niklas Koch

**Aufgabenstellung:** Der Vorgang zur Beschaffung geeigneter und genügender Texte für die Abdeckung aller Kategorien, welche für die Klassifikation und die Zusammenfassung genutzt werden können, muss dokumentiert werden.

**Zieldefinition:** Der Vorgang der Datenbeschaffung muss dokumentiert werden.

**Geplanter Zeitraum:** 09.07.23 - 23.07.23

**Voraussetzung:** Die Arbeitspakete unter dem Kapitel *Finden geeigneter Daten* müssen erledigt sein.

**Hilfsmittel:** LaTeX

**Risiken:** Eine unzureichende Dokumentation kann zu einem nicht nachvollziehbaren Vorgehen führen, wodurch die Transparenz des Projekts schwinden kann.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Die Arbeitspakete unter dem Kapitel *Finden geeigneter Daten* sind Voraussetzung für die Dokumentation.

**Abnahmebedingungen:** Die Dokumentation ist vollständig ausgearbeitet.

### 7.1.2 Dokumentation des Moduls zur Spracheingabe

**PSP-Code:** AP-712

**Verantwortlicher:** Niclas Cramer

**Aufgabenstellung:** Der Vorgang von der Recherche bis zur Implementierung der Funktion der Spracheingabe muss dokumentiert werden.

**Zieldefinition:** Der Vorgang der Spracheingabe muss dokumentiert werden.

**Geplanter Zeitraum:** 09.07.23 - 23.07.23

**Voraussetzung:** Die Recherche und Implementierung des Moduls zur Spracheingabe muss erfolgt sein.

**Hilfsmittel:** LaTeX

**Risiken:** Eine unzureichende Dokumentation kann zu einem nicht nachvollziehbaren Vorgehen führen, wodurch die Transparenz des Projekts schwinden kann.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Die Arbeitspakete im Zusammenhang mit der Spracheingabefunktionalität sind Voraussetzung für die Dokumentation.

**Abnahmebedingungen:** Die Dokumentation ist vollständig ausgearbeitet.



### 7.1.3 Dokumentation durch Kommentieren des Codes zur Spracheingabe

**PSP-Code:** AP-713

**Verantwortlicher:** Niclas Cramer

**Aufgabenstellung:** Der Vorgang zur Spracheingabe soll in der Code-Umgebung durch Kommentare nachvollziehbar dargestellt werden.

**Zieldefinition:** Der Code zur Spracheingabe muss kommentiert werden.

**Geplanter Zeitraum:** 09.07.23 - 23.07.23

**Voraussetzung:** Die Recherche und Implementierung des Moduls zur Spracheingabe muss erfolgt sein.

**Hilfsmittel:** Python, VSCode

**Risiken:** Eine unzureichende Kommentierung des Codes kann zu einem nicht nachvollziehbaren Vorgehen führen, wodurch die Transparenz des Projekts schwinden kann.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Die Arbeitspakete im Zusammenhang mit der Spracheingabefunktionalität sind Voraussetzung für die Dokumentation.

**Abnahmebedingungen:** Der Code zur Spracheingabe wurde vollständig und nachvollziehbar kommentiert.

#### **7.1.4 Dokumentation des Moduls zum Einlesen von Textdokumenten**

**PSP-Code:** AP-714

**Verantwortlicher:** Antoine Fuchs

**Aufgabenstellung:** Der Vorgang von der Recherche bis zur Implementierung der Funktion zum Einlesen von Textdokumenten muss dokumentiert werden.

**Zieldefinition:** Der Vorgang zum Einlesen von Textdokumenten muss dokumentiert werden.

**Geplanter Zeitraum:** 09.07.23 - 23.07.23

**Voraussetzung:** Die Recherche und Implementierung des Moduls zum Einlesen von Textdokumenten muss erfolgt sein.

**Hilfsmittel:** LaTeX

**Risiken:** Eine unzureichende Dokumentation kann zu einem nicht nachvollziehbaren Vorgehen führen, wodurch die Transparenz des Projekts schwinden kann.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Die Arbeitspakete im Zusammenhang mit dem Einlesen von Textdokumenten sind Voraussetzung für die Dokumentation.

**Abnahmebedingungen:** Die Dokumentation zum Einlesen von Textdokumenten ist vollständig ausgearbeitet.

### **7.1.5 Dokumentation durch Kommentieren des Codes zum Einlesen von Textdokumenten**

**PSP-Code:** AP-715

**Verantwortlicher:** Antoine Fuchs

**Aufgabenstellung:** Der Vorgang zum Einlesen von Textdokumente soll in der Code-Umgebung durch Kommentare nachvollziehbar dargestellt werden.

**Zieldefinition:** Der Code zum Einlesen von Textdokumenten muss kommentiert werden.

**Geplanter Zeitraum:** 09.07.23 - 23.07.23

**Voraussetzung:** Die Recherche und Implementierung des Moduls zum Einlesen von Textdokumenten muss erfolgt sein.

**Hilfsmittel:** Python, VSCode

**Risiken:** Eine unzureichende Kommentierung des Codes kann zu einem nicht nachvollziehbaren Vorgehen führen, wodurch die Transparenz des Projekts schwinden kann.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Die Arbeitspakete im Zusammenhang mit dem Einlesen von Textdokumenten sind Voraussetzung für die Kommentierung des Codes.

**Abnahmebedingungen:** Der Code zum Einlesen von Textdokumenten wurde vollständig und nachvollziehbar kommentiert.

## 7.2 Dokumentation der Klassifikation

### 7.2.1 Dokumentation der einzelnen Module zur Klassifikation

**PSP-Code:** AP-721

**Verantwortlicher:** Niklas Koch, Jasmina Pascanovic

**Aufgabenstellung:** Der Vorgang von der Recherche, über die Implementierung und Anbindung der einzelnen Module zur Klassifikation müssen dokumentiert werden.

**Zieldefinition:** Die einzelnen Module der Klassifikation müssen dokumentiert werden.

**Geplanter Zeitraum:** 09.07.23 - 23.07.23

**Voraussetzung:** Die Recherche und Implementierung der Module zur Klassifikation muss erfolgt sein.

**Hilfsmittel:** LaTeX

**Risiken:** Eine unzureichende Dokumentation kann zu einem nicht nachvollziehbaren Vorgehen führen, wodurch die Transparenz des Projekts schwinden kann.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Die Arbeitspakete im Zusammenhang mit der Klassifikation sind Voraussetzung für die Dokumentation.

**Abnahmebedingungen:** Die Dokumentation für die einzelnen Module der Klassifikation ist vollständig ausgearbeitet.

### 7.2.2 Dokumentation durch Kommentieren des Codes der Klassifikation

**PSP-Code:** AP-722

**Verantwortlicher:** Niklas Koch, Jasmina Pascanovic

**Aufgabenstellung:** Der Vorgang zur Klassifikation von Texten soll in der Code-Umgebung durch Kommentare nachvollziehbar dargestellt werden.

**Zieldefinition:** Der Code zur Klassifikation von Texten muss kommentiert werden.

**Geplanter Zeitraum:** 09.07.23 - 23.07.23

**Voraussetzung:** Die Recherche und Implementierung der Module zur Klassifikation muss erfolgt sein.

**Hilfsmittel:** Python, VSCode

**Risiken:** Eine unzureichende Kommentierung des Codes kann zu einem nicht nachvollziehbaren Vorgehen führen, wodurch die Transparenz des Projekts schwinden kann.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Die Arbeitspakete im Zusammenhang mit der Klassifikation sind Voraussetzung für die Kommentierung des Codes.

**Abnahmebedingungen:** Der Code zur Klassifikation von Texten wurde vollständig und nachvollziehbar kommentiert.

## 7.3 Dokumentation der Zusammenfassung

### 7.3.1 Dokumentation der einzelnen Module zur Zusammenfassung

**PSP-Code:** AP-731

**Verantwortlicher:** Niclas Cramer, Antoine Fuchs

**Aufgabenstellung:** Der Vorgang von der Recherche, über die Implementierung und Anbindung der einzelnen Module zur Zusammenfassung müssen dokumentiert werden.

**Zieldefinition:** Die einzelnen Module der Zusammenfassung müssen dokumentiert werden.

**Geplanter Zeitraum:** 09.07.23 - 23.07.23

**Voraussetzung:** Die Recherche und Implementierung der Module zur Zusammenfassung muss erfolgt sein.

**Hilfsmittel:** LaTeX

**Risiken:** Eine unzureichende Dokumentation kann zu einem nicht nachvollziehbaren Vorgehen führen, wodurch die Transparenz des Projekts schwinden kann.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Die Arbeitspakete im Zusammenhang mit der Zusammenfassung sind Voraussetzung für die Dokumentation.

**Abnahmebedingungen:** Die Dokumentation für die einzelnen Module der Zusammenfassung ist vollständig ausgearbeitet.

**7.3.2 Dokumentation durch Kommentieren des Codes zur Zusammenfassung**

**PSP-Code:** AP-732

**Verantwortlicher:** Niclas Cramer, Antoine Fuchs

**Aufgabenstellung:** Der Vorgang zur Zusammenfassung von Texten soll in der Code-Umgebung durch Kommentare nachvollziehbar dargestellt werden.

**Zieldefinition:** Der Code zur Zusammenfassung von Texten muss kommentiert werden.

**Geplanter Zeitraum:** 09.07.23 - 23.07.23

**Voraussetzung:** Die Recherche und Implementierung der Module zur Zusammenfassung muss erfolgt sein.

**Hilfsmittel:** Python, VSCode

**Risiken:** Eine unzureichende Kommentierung des Codes kann zu einem nicht nachvollziehbaren Vorgehen führen, wodurch die Transparenz des Projekts schwinden kann.

**Eventuelle Schnittstellen zu anderen Arbeitspaketen:** Die Arbeitspakete im Zusammenhang mit der Zusammenfassung sind Voraussetzung für die Kommentierung des Codes.

**Abnahmebedingungen:** Der Code zur Zusammenfassung von Texten wurde vollständig und nachvollziehbar kommentiert.