Masterarbeit

zur Erlangung des akademischen Grades
Master

Technische Hochschule Wildau
Fachbereich Ingenieur- und Naturwissenschaften
Studiengang Telematik (M. Eng.)

Thema (deutsch): Konzeption und prototypische Implementierung eines hybriden Algorithmus zur Emotionserkennung mit humanoiden Robotern

Thema (englisch): Conception and prototypical implementation of a hybrid algorithm for emotion recognition with humanoid robots

Autor/in: Tina Lütke

Seminargruppe: TM/18

Betreuer/in: Prof. Dr. rer. nat. Janett Mohnke

Zweitgutachter/in: Prof.Dipl.-Inf. Birgit Wilkes

Eingereicht am:

Spätestmögliche Abgabe: 29.10.2020

Masterarbeit

Antrag vom: 11.03.2020

Name:	Tina Lütke	Matrikel-Nr.:	50018611
Studiengang:	Telematik	Seminargruppe:	TM/18
Betreuende/r Hochschuldozent/in:	Prof. Dr. rer. nat. Janett Mohnke	Beginn der Arbeit:	29.03.2020
Zweitgutachter/in:	Prof.Dipl.-Inf. Birgit Wilkes	Abgabetermin:	29.10.2020
Themensteller (z.B. Betrieb, Institution, Behörde):	TH Wildau	Fachliche Betreuungsperson des Themenstellers:	Janett Mohnke
		Straße:	Hochschulring 2
		PLZ, Ort:	15745, Wildau

Kurzthema

Konzeption und prototypische Implementierung eines hybriden Algorithmus zur Emotionserkennung mit humanoiden Robotern

Kurzthema in Englisch

Conception and prototypical implementation of a hybrid algorithm for emotion recognition with humanoid robots

Zielstellung

Es soll ein hybrider Algorithmus zur Emotionserkennung, basierend auf schon existierenden Algorithmen und Open-Source Software, konzipiert und prototypisch für den Roboter ROS-E implementiert werden.

Inhaltliche Anforderungen / Teilaufgaben

- Recherche zum Thema Emotionen
- Recherche zum Thema Algorithmen zur Emotionserkennung
- Anforderungsanalyse
- Konzeption für den Algorithmus
- prototypische Implementierung für den Roboter ROS-E
- Entwickler- und Anwenderdokumentation

Sprache der Arbeit

Deutsch

Konsultationen erfolgen nach Vereinbarung mit dem betreuenden Hochschullehrer.

Prof. Dr. rer. nat. Janett Mohnke	Prof.Dipl.-Inf. Birgit Wilkes	Tina Lütke	<i>genehmigt</i>
(Hochschuldozent/in)	(Zweitgutachter/in)	(Student/in)	(Prüfungsausschuss)

Eingang der Abschlussarbeit _____

Sperrvermerk ☐ Ja ☒ Nein

BIBLIOGRAFISCHE BESCHREIBUNG

Lüthe, Tina

Konzeption und prototypische Implementierung eines hybriden Algorithmus zur Emotionserkennung mit humanoiden Robotern.

Masterarbeit, Technische Hochschule Wildau 2020, 139 Seiten, 34 Abbildungen, 3 Listings, 3 Tabellen, 4 Anlagen

Ziel

Das Ziel dieser Masterarbeit ist die Konzeption und prototypische Implementierung einer hybriden Emotionserkennung für den Tischroboter ROS-E. Die Emotionserkennung soll eine Emotion aus mindestens zwei verschiedenen Eingaben bestimmen. Die Grundlage des Konzepts sollen Dienste, Open-Source-Projekte oder bereits umgesetzte Ansätze bilden.

Inhalt

Zuerst wird eine Einleitung in die Masterarbeit gegeben, in welcher der Hintergrund für diese Idee, das Ziel und die Abgrenzung dieser Thesis genauer erläutert werden.

Anschließend wird das Ziel dieser Arbeit konkretisiert, indem Einsatzszenarien, die Zielgruppe und Forschungsfragen verfasst werden.

Im Anschluss daran werden sowohl Grundlagen sozialer Aspekte, als auch die technischen Grundlagen aufgeführt. Die Grundlagen umfassen die Definition von Emotionen, die Darstellung der Kommunikation zwischen Mensch und Roboter, einen Einblick in den Roboter ROS-E, die Erklärung der Lizenzen von Open-Source-Software und die Erläuterung von benötigten Teilgebieten der künstlichen Intelligenz.

Daraufhin werden die verschiedenen Dienste, Open-Source-Software und Ansätze zur Emotionserkennung vorgestellt.

Basierend auf diesen Grundlagen wird ein Konzept zur hybriden Emotionserkennung entwickelt, welches anschließend prototypisch umgesetzt wird.

Auf der beigelegten CD-ROM befinden sich das PDF der Masterarbeit, sowie die ausführbaren Dateien, eine Anwenderdokumentation und Beispiel-Daten.

ABSTRACT

Diese Masterarbeit befasst sich mit der Recherche zum Thema Emotionserkennung, der Konzeption einer hybriden Emotionserkennung und der prototypischen Implementierung des Konzepts für den Tischroboter ROS-E.

Die Recherche zeigt, dass die Emotionserkennung auf Basis von verschiedenen Eingaben erfolgen kann. Es wurde sich auf die drei Eingabeströme Bild, Ton und Text konzentriert, da diese praktikabel für die Implementierung auf einen humanoiden Roboter sind. Zu jedem dieser Eingabeströme wurden Dienste, Open-Source-Software und Ansätze untersucht.

Aufgrund dieser Erkenntnisse wurde eine Anwendung entwickelt, welche eine hybride Emotionserkennung durch die Eingabe eines Textes und einer Bild-Datei durchführt. Hierzu wurden die Dienste zur Emotionserkennung von Microsoft Azure und IBM Watson genutzt. Die Anwendung wurde in Python implementiert.

This master thesis deals with the research of emotion recognition, the conception of a hybrid emotion recognition and and prototypical implementation of the concept for the robot ROS-E.

The research has shown, that emotion recognition can be performed based on different types of inputs. It has been focused on the input streams image, audio and text, since they are the most practical for the implementation on a humanoid robot. Different services, open source software and approaches have been researched for each of these input streams.

Based on these knowledges an application was developed, which executes hybrid emotion recognition through the input of a text and an image. The application implements the emotion recognition services of Microsoft Azure and IBM Watson. Furthermore the application was implemented in Python.

EIDESSTATTLICHE ERKLÄRUNG

Ich erkläre hiermit, dass ich die vorliegende Abschlussarbeit selbstständig angefertigt und nur die angegebenen Hilfsmittel und Quellen verwendet habe.

(Ort, Datum)

(Unterschrift)

INHALTSVERZEICHNIS

INHALTSVERZEICHNIS	IV
DANKSAGUNG	VII
HINWEISE ZUM LESEN DER ARBEIT	VIII
1. Einleitung.....	1
2. Konkretisierung der Thematik	4
2.1.Einsatzszenarien.....	4
2.2.Personae.....	7
2.3.Forschungsfragen	13
3. Grundlagen sozialer Aspekte.....	15
3.1.Emotionen	15
3.2.Kommunikation zwischen Mensch und Roboter	25
4. Technische Grundlagen	30
4.1.ROS-E	30
4.2. Lizenzen	33
4.3.Künstliche Intelligenz.....	35
4.3.1.Machine Learning	35
4.3.2.Deep Learning	35
4.3.3.Künstliches Neuronales Netz (KNN)	36
4.3.4.Convolutional Neural Network (CNN).....	37
4.3.5.Mehrlagiges Perzeptron (MLP).....	39
4.3.6.Recurrent Neural Network (RNN)	39
4.3.7.Long Short-Term Memory (LSTM).....	40
4.3.8.Deep Neural Network (DNN)	40
5. Stand der Technik	41

5.1.Die drei Grundpfeiler der Emotionserkennung.....	41
5.2. Open-Source-Software, -Projekte und Dienste	43
5.2.1.Bild.....	43
5.2.2.Ton.....	47
5.2.3.Text	48
5.2.4.Multimodal	49
5.3.Algorithmen	52
5.3.1.Bild.....	52
5.3.2.Ton.....	59
5.3.3.Text.....	61
6. Konzept.....	67
6.1.Abgrenzung.....	67
6.2.Konzept 1	70
6.3.Konzept 2.....	74
7. Prototypische Implementierung	77
7.1.Implementierung des ersten Ansatzes.....	77
7.2.Implementierung des zweiten Ansatzes.....	78
7.2.1.Einsatz der Anwendung.....	78
7.2.2.Struktur der Anwendung	79
7.2.3.Klassen.....	80
7.2.4.Benutzeroberfläche	83
7.2.5.Aufgetretene Probleme.....	85
7.2.6.Test	86
8. Ausblick	88
8.1.Implementierung auf ROS-E.....	88
8.2.Nutzung von Live-Daten	88

8.3.Nutzung von Video-Daten	89
8.4.Erweiterung der Anwendung um den Eingabestrom Ton	89
8.5.Einbeziehen zusätzlicher Komponenten	90
8.6.Nutzung eines KNNs.....	90
9. Zusammenfassung und Fazit.....	92
ANHANG A - CD-ROM.....	VIII
ANHANG B - MIT-Lizenz.....	IX
ANHANG C - Apache-2.0-Lizenz.....	X
ANHANG D - AGPL-3.0-Lizenz.....	XI
GLOSSAR.....	XII
ABKÜRZUNGSVERZEICHNIS	XVII
ABBILDUNGSVERZEICHNIS	XIX
TABELLENVERZEICHNIS	XXII
LITERATURVERZEICHNIS	XXIII

DANKSAGUNG

Zunächst möchte ich mich an dieser Stelle bei all denjenigen bedanken, welche mich nicht nur während der Anfertigung dieser Masterarbeit, sondern während meines gesamten Studiums unterstützt und motiviert haben.

Ganz besonderer Dank gilt meinen Betreuern Prof. Dr. rer. nat. Janett Mohnke und Prof. Dipl-Inf. Birgit Wilkes, welche mir nicht nur vor und während meiner Masterarbeit, sondern auch in meinem gesamten Studium, mit hilfreichen Hinweisen und fachlichem Rat beiseite standen.

Zu guter Letzt möchte ich mich bei meiner Familie, meinen Freunden und meinem Freund bedanken, welche stundenlang Korrektur gelesen, Programme getestet und mir die nötige Ablenkung gegeben haben. Danke für Eure aufmunternden und motivierenden Worte und auch eure positive Art, mich emotional zu unterstützen.

Des Weiteren möchte ich mich für Eure Unterstützung in meinem gesamten Studium bedanken, welches nicht immer leicht war. Auch hier möchte ich Euch danken, dass ihr immer an mich geglaubt habt.

HINWEISE ZUM LESEN DER ARBEIT

Nachfolgend werden die formellen Aspekte dieser Bachelorarbeit genannt und genauer erläutert.

Die Quellenangaben in Texten werden am Ende der jeweiligen Abschnitte in eckigen Klammern und der entsprechenden Nummer auf eine Quelle aus dem Quellenverzeichnis dargestellt. Bei Abbildungen und Listings werden die Quellen in gleicher Form am Ende der Bildbeschreibung angegeben. Selbst erstellte und entworfene Abbildungen und Listings erhalten keinen Quellbezug.

Zitate werden am Anfang und am Ende des Zitats mit Anführungszeichen gekennzeichnet. Auslassungen eines Zitats werden mit drei Punkten in eckigen Klammern [...] kenntlich gemacht. Nach einem Zitat wird die Quelle, wie oben beschrieben, angegeben.

Eigennamen oder Firmen werden, falls benötigt, durch eine Fußnote kurz erklärt und, falls existent, auf die eigene Webseite verwiesen.

Fachspezifische, nicht im Duden aufgeführte Abkürzungen werden im Abkürzungsverzeichnis angegeben und dort ausgeschrieben.

Wörter, deren Bedeutung unklar ist, werden im Glossar erwähnt. Dort werden Erklärungen zu diesen Wörtern angegeben. Die im Glossar stehenden Wörter werden durch kursive Schrift kenntlich gemacht.

Bezeichnungen von Variablen, Dateien, Ordnern oder Ähnlichem werden in Anführungszeichen geschrieben.

Um wesentliche Informationen innerhalb eines Abschnitts oder Kapitels hervorzuheben, werden wichtige Schlagworte oder Wortgruppen fett markiert.

1. Einleitung

"At bottom, robotics is about us. It is the discipline of emulating our lives,
of wondering how we work." [18]

- Roderic A. Grupen¹

Dass humanoide Roboter in unserem alltäglichen Leben auftreten, ist schon lange keine Seltenheit mehr. Wir finden sie in verschiedenen Lebenssituationen; im Supermarkt, in der Bibliothek oder in Pflegeeinrichtungen. Und obwohl wir für humanoide Roboter eine gewisse Empathie empfinden, die teilweise mit der für den Menschen empfundene Empathie gleichgestellt werden kann, gestaltet sich die Kommunikation oft noch sehr kompliziert. [1]

Offensichtliche Gründe hierfür sind zum Beispiel die langsamen Reaktionszeiten des Roboters, oder dass diese aufgrund ihrer noch sehr roboterhaften Stimme noch nicht menschlich genug wirken. Einer der nicht so offensichtlichen Gründe ist, dass die Roboter unsere Mimik und Gestik noch nicht gut genug erkennen und dementsprechend in das Gespräch mit einbeziehen können. Eine große Verbesserung kann hier das Einbeziehen von Emotionen mit sich bringen. Firmen wie Microsoft², IBM³ und Amazon⁴ haben dies bereits erkannt und bieten Algorithmen zur Emotionserkennung an. Amazon setzt hierbei auf die "klassische" Emotionserkennung durch die Gesichtszüge einer Person, IBM hingegen setzt auf Emotionserkennung durch Text. [2] [3] [4] [28] [30]

1 Professor für Informatik und Direktor des Labors für *Perceptual Robotics* an der University of Massachusetts [15]

2 Microsoft: <https://www.microsoft.com/>

3 IBM: <https://www.ibm.com/>

4 Amazon: <https://www.amazon.de/>

Einer neuen Studie¹ zufolge sind Emotionen sehr viel komplexer als meist angenommen. Emotionen sollen nicht nur anhand des Aussehens von Personen erkannt werden können. Somit kann eine Person zum Beispiel ein böses Gesicht machen, was aber nicht zu bedeuten hat, dass die Person gerade Wut fühlt. Emotionen sind subjektiv. Menschen schreien, weinen, lachen oder planen ihre Rache, wenn sie Wut empfinden. Die Reaktionen unterscheiden sich von Mensch zu Mensch und von Situation zu Situation. [2]

Aus diesen Gründen soll sich diese Thesis mit einer hybriden Lösung zur Emotionserkennung beschäftigen, welche sich aus mehreren Faktoren, wie Bild-, Audio- oder Texterkennung zusammensetzen soll.

Ziel der Thesis ist die Recherche zu Emotionen und entsprechenden Algorithmen zur Erkennung dieser, die Konzeption eines hybriden Algorithmus zur Erkennung von Emotionen basierend auf den Ergebnissen der Recherche und die prototypische Implementierung des Konzepts. Das Konzept soll für den Roboter ROS-E implementiert werden. ROS-E soll somit die Möglichkeit erhalten, dass sich die Kommunikation zwischen Mensch und Roboter durch die Emotionserkennung natürlicher verhält. Der Fokus der zu verwendenden Dienste zur Umsetzung des Algorithmus liegt hier auf nicht kommerziellen Diensten. Dies gilt ebenso für Frameworks und Ähnlichem, was in dieser Thesis genutzt wird.

Die Recherche dieser Thesis wird sich mit Anwendungsszenarien beschäftigen und wie diese mithilfe eines Algorithmus gelöst worden sind oder gelöst werden können. Hierbei werden ethische Aspekte außen vor gelassen, da sich diese Arbeit damit beschäftigt, wie gut ein Roboter überhaupt Emotionen erkennen kann. Ebenso wird sich die Thesis nicht mit dem perfekten Algorithmus für alle Roboter und für alle Anwendungsfälle beschäftigen. Stattdessen sollen vorhandene Lösungen analysiert und *Open-Source*-Software genutzt werden, um ein passendes Konzept für den Roboter ROS-E zu entwickeln. Das Konzept wird prototypisch umgesetzt.

Wenn der Algorithmus auf einem Roboter implementiert worden ist, kann er von jeder Alters- und Personengruppe genutzt werden, die diesen Roboter nutzt. Hierbei kann es sich

1 Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements

um einen Lern-Roboter für Kinder handeln oder auch um einen Service-Roboter im Supermarkt.

Im folgenden Teil dieser Arbeit wird die gesamte Thematik konkretisiert. Hierbei werden Einsatzszenarien beschrieben, in denen die Emotionserkennung genutzt werden kann. Des Weiteren wird die Zielgruppe anhand von Personae genauer eingegrenzt. Anschließend werden Forschungsfragen formuliert, welche in dieser Thesis beantwortet werden.

Im dritten Kapitel werden die sozialen Grundlagen behandelt, die für diese Thesis von Bedeutung sind. Dazu gehören die Erläuterung von Emotionen und die gängige und aktuelle Kommunikation zwischen Mensch und Roboter.

Im Anschluss daran werden die technischen Grundlagen erläutert. Hierzu wird zuerst der Roboter ROS-E vorgestellt. Des Weiteren wird auf Lizenzen von *Open-Source*-Software eingegangen. Schließlich werden benötigte Begriffe der künstlichen Intelligenz erläutert, um ein besseres Verständnis für das folgende Kapitel zu bekommen.

Im nächsten Kapitel wird auf den Stand der Technik eingegangen. Hierbei werden die Komponenten genannt und erläutert, mit denen Emotionen erkannt werden können. Des Weiteren wird auf verschiedene Dienste, *Open-Source*-Software und Ansätze zur Emotionserkennung eingegangen.

Im sechsten Kapitel wird aufgrund der gewonnenen Erkenntnisse zwei Konzepte erstellt. Hierfür werden die im vorigen Kapitel beschriebenen Dienst, *Open-Source*-Software und Ansätze genutzt.

Im nächsten Kapitel wird die Implementierung der Konzepte beschrieben. Hierzu werden auf den Einsatz, die Struktur, die Klassen und die Benutzeroberfläche eingegangen. Des Weiteren werden aufgetretene Probleme und die Testphase erläutert.

Zuletzt wird auf den Ausblick dieser Arbeit eingegangen. Im Anschluss wird eine Zusammenfassung gegeben und die Forschungsfragen beantwortet.

2. Konkretisierung der Thematik

In diesem Kapitel wird genauer erläutert, für welche Szenarien diese Thesis interessant sein bzw. wo Emotionserkennung genutzt werden kann. Weiterhin soll klar werden, für wen das Ergebnis dieser Thesis interessant ist und welchen Rahmen diese Thesis umfassen soll.

2.1. Einsatzszenarien

Für die Emotionserkennung lassen sich eine Reihe Einsatzszenarien finden. Hierbei gibt es einige, die den Menschen das Leben erleichtern, andere wiederum sind nützlich und manche nur Spielereien. Auch hier werden sich Szenarien finden, mit denen viele von uns aus Gründen der Privatsphäre nicht einverstanden wären. Abgesehen davon ist es interessant zu sehen, inwiefern die Emotionserkennung unser Leben mit Technik verändern kann. Aus diesem Grund werden hier soziale und ethische Faktoren nicht beachtet.

Viele Einsatzszenarien beruhen darauf, dass mithilfe von Emotionserkennung ein besseres Nutzererlebnis geschaffen werden kann. Vorstellbar wäre eine Emotionserkennung während des **Online-Shoppings**. Durch die Emotionserkennung kann ohne jegliche Bewertung durch den Nutzer ein Feedback für den Anbieter erstellt werden. Dieses kann sich auf bestimmte Produkte, den Shop-Aufbau, die Bestellabwicklung oder das Kontaktformular beziehen. Der Anbieter bekommt auf diese Art und Weise eine Menge an Informationen und kann Schwachstellen ausbessern und Gelungenes hervorheben. [5]

Gleiches kann für andere Online-Angebote genutzt werden, zum Beispiel für **Online-Unterricht, -Vorlesungen oder -Kurse**. Der dahinterstehende Verantwortliche kann seinen Inhalt auf die Nutzer abstimmen. Sind Gesagtes und die enthaltenen Informationen verständlich? Hören alle zu? Langweilen sich einige? Dies kann zur Verbesserung der Vortragsweise des Verantwortlichen oder zur inhaltlichen Verbesserung genutzt werden. [5]

Ein weiteres ähnliches Szenario ist die Emotionserkennung zur Bewertung von **TV-Sendungen, -Sendern oder Streaming-Diensten**. Hier kann die Emotionserkennung genauso genutzt werden wie bei den vorherigen Szenarien, um denselben Zweck zu erfüllen. [5]

Ein letztes ähnliches Szenario, das zum selben Zweck genutzt werden kann, ist die Verbesserung von **Virtual-Reality** Erlebnissen, sowohl für PC- oder Konsolenspiele, Wohnungsbesichtigungen oder Ähnlichem. [5]

An öffentlichen Orten wäre es denkbar, mithilfe von Emotionserkennung böse Absichten erkennen zu können, um zusätzliche **Sicherheit** zu gewähren. Das kann ein Taschendiebstahl in einem Einkaufszentrum, ein Terroranschlag auf dem Flughafen oder eine mutwillige Zerstörung auf dem Sportplatz sein. [5]

Ein spannendes Einsatzszenario stellt die Emotionserkennung für **Autisten** dar. Diese besitzen nicht die Fähigkeit, Emotionen zu erkennen oder diese selbst auszudrücken und haben somit Schwierigkeiten, einen sozialen Umgang mit anderen Menschen zu pflegen. Hier wäre es denkbar, die Emotionserkennung als *Wearable* oder integriert in einer *Smart Watch* zu realisieren. Es existiert bereits eine Applikation, die Autisten dabei helfen soll, Emotionen erlernen zu können. Die Applikation "Zirkus Empathie" ist ähnlich wie eine Applikation zum Sprachenlernen entwickelt worden. Genauso wäre es denkbar, dass ein Autist eine Applikation zur Verfügung hat, mit der er diese erlernten Emotionen an sich selbst ausprobieren kann. Die Applikation könnte dann seine eigene dargestellte Emotion erkennen. Somit kann der Autist überprüfen, ob dies die Emotion ist, welche er darstellen wollte. [5] [31]

Ein letztes Einsatzszenario stellen **humanoide Roboter** dar. Während wir diese vor ein paar Jahren nur in den Medien gesehen haben, finden wir sie nun immer häufiger in unserem Alltag. Saturn hat vor einigen Jahren den Roboter

Abb. 1: Roboter Paul im Elektronikfachhandel Saturn [7]

Paul eingesetzt, um Kunden zum gewünschten Produkt im Geschäft zu führen. Paul ist in Abbildung 1 zu erkennen. Auch für ein wenig Smalltalk kann dieser genutzt werden. Ein anderer Roboter (siehe Abbildung 2), der kürzlich in einem Conrad in Berlin eingesetzt wurde, kann übliche Fragen der Kunden beantworten und einen Kauf abwickeln. In einem Pflegeheim in Siegen wird der Roboter Paula als Beschäftigung eingesetzt, wie in Abbildung 3 zu sehen ist. Paula kann Musik spielen, Spiele spielen, zwischendurch Witze erzählen und damit die Pfleger zum Teil entlasten. [6] [7] [8]

Abb. 2: Roboter im Elektronikfachhandel Conrad [6]

Abb. 3: Roboter Paula aus einem Pflegeheim in Siegen [8]

Von humanoiden Robotern erwartet man, dass diese dem Menschen sehr ähnlich sind, ähnlich denken, reden oder reagieren. Meistens wird man genau dann enttäuscht, da die Technik keinem realen Menschen entspricht. Wenn der humanoide Roboter aber Emotionen erkennen und entsprechend darauf reagieren kann, wirkt er realer. Hierbei kann es sich um die verschiedensten Szenarien handeln. Sei es der Roboter als zusätzliche Hilfe im Elektrofachhandel oder im Supermarkt, als Pflegekraft im Altersheim oder als Lernroboter für Kinder. Jede Altersgruppe in jeder Situation fühlt sich besser verstanden, wenn auf die eigenen Emotionen Acht genommen wird.

All die genannten Einsatzszenarien wären vorstellbar. Das Hauptaugenmerk wird in dieser Thesis jedoch auf die **Emotionserkennung mit humanoiden Robotern** gelegt. Die Zielgruppe für dieses Einsatzszenario ist sehr breit gefächert. Um die

Bedürfnisse dieser besser zu verstehen und darauf eingehen zu können, werden zunächst die Zielgruppen mithilfe von Personae genauer bestimmt.

2.2. Personae

In diesem Kapitel werden Personae erstellt, um die Zielgruppe genauer definieren zu können. Die Zielgruppe für diese Thesis setzt sich aus den verschiedensten Personen- und Altersgruppen zusammen. Aufgrund dessen haben diese Gruppen auch verschiedene Bedürfnisse und Erwartungen an die Emotionserkennung mit humanoiden Robotern.

Die Personae werden aufgrund dieser Thesis auf humanoide Roboter bezogen. Für diese Thesis sind zwei Personengruppen von Bedeutung, welche in den folgenden Personae genauer beschrieben werden. Eine der Personengruppen sind die **Anwender**, welche sehr vielfältig sein können. Dabei können diese, wie auch in den Personae erkennbar, Schüler, Erwachsene mittleren Alters oder auch Rentner sein. Jedoch haben sie eines gemeinsam. Die Anwender haben ein starkes Interesse an (neuester) Technik, haben aber oftmals kein tiefgründiges Wissen. Wenn die Anwender Robotern gegenüberstehen, zeigen sie großes Interesse und sind sehr aufgeschlossen. Trotz der Freude darüber, die Roboter ausprobieren zu können, sind sie oft auch ein wenig enttäuscht, wenn die Roboter nicht wie in ihren Lieblingsfilmen agieren und reagieren, dafür aber eher wie Maschinen wirken. Sie wünschen sich, dass genau das verbessert wird.

Lisa Maier

Alter: 8 Jahre

Beruf: Grundschülerin

Technisches Interesse:

- Sie besitzt kein tiefgründiges Wissen.
- Sie hat Interesse an alltäglicher Technik, wie Fernseher und Tablet.
- Ihr Interesse wird auch durch die Eltern geweckt, die in der IT-Branche arbeiten.
- Sie nutzt den in der Schule angebotenen Lernroboter zum Verbessern der eigenen Fähigkeiten beim Lesen, Rechnen und in Englisch.

Abb. 4: Grundschülerin Lisa Maier
[10]

Erwartungen und Wünsche:

- Sie hat die Erwartung, dass der Roboter lustig ist und sie versteht.
- Sie kann mit ihm umgehen, wie sie es auch mit einem Menschen tun würde.

Probleme und Ängste:

- Sie hat Angst, dass der Roboter sie nicht versteht.
- Der Roboter wirkt zu unmenschlich und deswegen könnte er ihr Angst machen.

Dirk Kaufmann

Alter: 36 Jahre

Beruf: Malermeister

Technisches Interesse:

- Er hat ein starkes eigenes Interesse an Technik.
- Er hat kein tiefgründiges Wissen.
- Er bezieht sein Wissen hauptsächlich aus Nachrichten, Artikeln und Zeitschriften.
- Er kennt Roboter aus Elektronikmärkten, den Nachrichten und aus Artikeln.

Abb. 5: Malermeister Dirk Kaufmann [11]

Erwartungen und Wünsche:

- Aufgrund des eigenen Wissens kennt er die Schwächen der Roboter, weshalb er keine hohen Erwartungen hat.
- Wünschen würde er sich, dass man eine nahezu Mensch-zu-Mensch-Kommunikation mit einem Roboter führen kann.

Probleme und Ängste:

- Die Roboter wirken zu kompliziert und nicht menschlich genug.

Martin Müller

Alter: 82 Jahre

Beruf: Rentner, ehemals Bauleiter

Technisches Interesse:

- Er hat eigenes Interesse an alltäglicher Technik, wie Smartphones und Tablets.
- Sein Wissen bezieht er hauptsächlich aus den Nachrichten, aus Zeitschriften oder von Verwandten.
- Er überlegt, in ein Altersheim mit einem humanoidem Roboter zu gehen oder einen Roboter als Kumpel für zu Hause zu kaufen, um Hilfe im Alltag zu bekommen und nicht einsam zu sein.
- Interesse an den humanoiden Robotern wurde durch Kinder und Enkelkinder geweckt

Abb. 6: Rentner Martin Müller [12]

Erwartungen und Wünsche:

- Er wünscht sich eine leichte Bedienung für den Roboter oder, dass kaum Bedienung nötig ist.

Probleme und Ängste:

- Er hat Angst, dass Roboter (große) Fehler machen können.
- Er hat Angst, dass Roboter ihn nicht verstehen können, oder andersherum, zum Beispiel durch eine Sprachein- oder Sprachausgabe.
- Er hofft, dass die Bedienung des Roboters nicht zu aufwendig oder kompliziert ist.

Die zweite Personengruppe sind die **Entwickler**. Die Entwickler arbeiten in verschiedenen Positionen, aber jeder von ihnen arbeitet an der Entwicklung von Software oder Hardware von Robotern, in einer höheren Stelle eines solchen Unternehmens oder als Produkttester. Die Motivation zur Emotionserkennung von

Robotern können verschiedener Natur sein. Einerseits kann diese komplett aus Eigeninteresse bestehen, aus dem Interesse für das eigene Unternehmen oder der Projektleiter oder eine andere übergeordnete Position die Recherche oder Verbesserung des Roboters als Aufgabe gestellt hat. Die Entwickler haben ihr Interesse vor allem an der Funktionsweise einer Emotionserkennung und wie sie diese mit in ihre Projekte einbinden können.

Laura Schmidt

Alter: 31 Jahre

Beruf: Softwareentwicklerin für ein Robotics-Unternehmen

Technisches Interesse:

- Sie besitzt ein starkes eigenes Interesse an Technik.
- Sie interessiert sich vor allem für die Bereiche Robotik, Automatisierungstechnik und Smart Home.
- Hauptsächlich betreibt sie Informationsbeschaffung für die Arbeit und aus persönlichem Interesse. Informationen bezieht sie aus Artikeln, Zeitschriften oder direkt von anderen Unternehmen.

Abb. 7: Softwareentwicklerin Laura Schmidt [13]

Erwartungen und Wünsche:

- Sie hat realistische Erwartungen und Wünsche an Robotern.
- Sie hofft, dass Roboter sich stetig verbessern, indem sie menschlicher gemacht werden oder menschlicher wirken sollen.
- Sie erwartet, dass die Kommunikation zwischen Mensch und Roboter verbessert wird.

Probleme und Ängste:

- Sie weiß aus erster Hand, dass das Preis-Leistungs-Verhältnis stimmen muss.
- Sie hofft, dass die Roboter nicht zu sehr vermenschlicht werden, da dies eher unangenehm wirken kann.

Christian Schneider

Alter: 35 Jahre

Beruf: Systemtester für ein Robotics-Unternehmen

Technisches Interesse:

- Er hat ein starkes eigenes Interesse an Technik und Robotern.
- Er besitzt kein tiefgründiges Wissen.
- Er bezieht sein Wissen hauptsächlich aus Artikeln, Zeitschriften und von der Arbeit.
- Er kennt Roboter von der Arbeit und kennt genauso ihre Stärken und Schwächen.

Abb. 8: Systemtester Christian Schneider [14]

Erwartungen und Wünsche:

- Er wünscht sich, dass Roboter mehr in den Alltag integriert werden.
- Er wünscht sich, dass Roboter sich menschlicher verhalten werden.

Probleme und Ängste:

- Er hat ein Problem damit, dass die Roboter ein zu unmenschliches Verhalten in Kommunikation und Interaktion aufzeigen.

Aus den definierten Personae und der damit verbundenen Zielgruppe ergeben sich die einzelnen Bedürfnisse der Personengruppen. Diese sollen in der gesamten

Thesis beachtet werden. Im nächsten Kapitel werden aufgrund der Personae Forschungsfragen aufgestellt.

2.3. Forschungsfragen

Die Forschungsfragen entstehen aus den Bedürfnissen und Erwartungen der Zielgruppe und sollen die Thesis zu einem Ergebnis leiten. Somit stellen sie einen groben roten Faden dieser Arbeit dar und grenzen diese ab. Die folgenden Forschungsfragen sollen im Laufe der Thesis beantwortet und im Fazit wieder aufgegriffen werden.

- **In welcher Form können sich Emotionen zeigen?**
 - Es soll herausgefunden werden, wie Emotionen physisch und psychisch bei einem Menschen erkennbar werden. Dazu sollen die psychologischen Hintergründe von Emotionen erforscht werden.
- **Wie wirken Roboter auf den Menschen aus heutiger Sicht?**
 - Es soll gezeigt werden, wie Menschen die Kommunikation und Interaktion mit humanoiden Robotern empfinden. Das macht erkenntlich, welche Eigenschaften an der Interaktion zwischen Mensch und Roboter erhalten werden sollten, welche verbessert werden können und welche Rolle Emotionen dabei spielen können. Ebenso können hiermit die Bedürfnisse der Zielgruppe noch genauer bestimmt werden.
- **Wie können Emotionen durch Computer erfasst werden? / Welche Daten können zur Auswertung von Emotionen genutzt werden?**
 - Es soll recherchiert werden, durch welche Eingabeströme oder Sensoren Emotionen erkannt werden können. Dies kann zum Beispiel bedeuten, dass Computer mithilfe eines Mikrofons die Stimme des Menschen analysieren

und daraus die Emotionen erkennen können. Es sollen mehrere Eingabeströme recherchiert werden.

- **Welche Algorithmen und/oder Dienste gibt es, die die Daten von Emotionen auswerten können?**

- Es sollen Algorithmen und/oder Dienste recherchiert werden, die die vorher recherchierten Eingabeströme nutzen, um die Emotion daraus zu bestimmen. Diese Algorithmen und Dienste sollen *Open Source* sein, um diese im nächsten Schritt nutzen zu können.

- **Welches hybride Konzept, das auf Algorithmen und/oder *Open Source* Diensten basiert, ermöglicht es einem humanoiden Roboter Emotionen zu erkennen?**

- Es soll ein Konzept für den Roboter ROS-E entwickelt werden, welches auf den zuvor recherchierten Algorithmen und/oder Diensten basiert und anschließend prototypisch implementiert werden. Dabei ist es wichtig, dass ein hybrides Konzept entworfen wird, also mehrere Faktoren zur Emotionserkennung genutzt werden.

Mit diesem Kapitel wurden die Möglichkeiten der Thematik aufgezeigt, die Zielgruppe konkretisiert, ein genauerer roter Faden und die zu erreichenden Ziele beschrieben. Auf Grundlage dieses Kapitels wird die Arbeit fortgeführt.

3. Grundlagen sozialer Aspekte

In diesem Kapitel werden die Grundlagen erläutert, welche die sozialen Aspekte dieser Arbeit umfassen. Hierbei wird vor allem auf die Emotionen, deren psychologischen Hintergrund und deren Ausdrucksweise, sowie die Rolle dieser bei humanoiden Robotern. Des Weiteren wird auf die aktuelle Kommunikation und Interaktion zwischen Mensch und Roboter eingegangen und inwiefern Emotionen diese verbessern können.

3.1. Emotionen

"Emotionen berühren uns, erregen uns und sie bewegen uns in eine bestimmte Richtung." [20, S. 165]

So wurden die Emotionen in dem Buch "Allgemeine Psychologie: Motivation und Emotion" zusammengefasst beschrieben. Denn Emotionen beschreiben Gefühle, Stimmungen und *Affekte*, die durch bestimmte Einflüsse ausgelöst werden können, zum Beispiel durch Wahrnehmungen, Gedanken oder Erinnerungen. Emotionen lassen sich an einem veränderten, subjektiven psychischen oder physischen Zustand erkennen. Diese *affektiven* Reaktionen werden unwillkürlich ausgelöst und sind zeitlich begrenzt. Dass Emotionen uns in eine bestimmte Richtung bewegen können, kann am Beispiel der Emotion Angst sehr gut erkannt werden. Angst motiviert Menschen zur Vermeidung von bestimmten Situationen oder Objekten und Ekel motiviert Zurückweisung. Emotionen senden ebenfalls Signale aus. Somit signalisieren positive Emotionen, wie zum Beispiel Freude Erfolge und negative Emotionen signalisieren Misserfolge. Damit beeinflussen sie uns in unseren Bedürfnissen und bei der Erreichung unserer Ziele. [9] [20, S. 165 ff.]

Emotionen besitzen vier Kennzeichen, um diese als Emotionen erkennen zu können. Ihre Affektivität, Objektgerichtetheit, Unwillkürlichkeit und ihre begrenzte Dauer. [20, S. 165 f.]

Die **Affektivität** von Emotionen beschreibt, dass Empfindungen oder deren Ursachen uns nicht zwingend bewusst sein müssen. Allerdings können wir uns unserem Empfinden oder deren Ursachen bewusst werden, sobald wir darauf aufmerksam werden. Die Affektivität wird auch Gefühlscharakter genannt. [20, S. 166]

Die **Objektgerichtetheit** bedeutet, dass Emotionen sich immer auf etwas (ein Objekt) beziehen. Dabei müssen die Objekte nicht tatsächlich existieren, sie können auch der Vorstellungskraft entsprungen sein, in der Vergangenheit oder in der Zukunft liegen. Die Objektgerichtetheit wird auch Intentionalität genannt. [20, S. 166]

Die **Unwillkürlichkeit** beschreibt das automatische Auslösen von Reaktionen. Wir können uns diesen Reaktionen, die auf die Emotionen folgen, nicht entziehen. Jedoch können wir Situationen und Informationen aufsuchen, damit solche Emotionen nicht empfunden werden. [20, S. 166]

Das letzte Kennzeichen einer Emotion ist ihre **zeitlich begrenzte Dauer**. Diese sagt aus, wie auch vorher schon erwähnt, dass Emotionen nur über eine bestimmte Dauer vorhanden sind. [20, S. 166]

Mit diesen Kennzeichen lassen sich **Emotionen von** Stimmungslagen, emotionalen Dispositionen (Temperamente) und **Charakterzügen abgrenzen**. Stimmungslagen haben keine Bezugsobjekte und dauern oft sehr viel länger an, als Emotionen. Temperamente hingegen zählen zu einer Eigenschaft der Persönlichkeit und ziehen sich damit über eine längere Dauer als Emotionen. Auch diese müssen sich nicht zwingend auf ein Objekt beziehen. [20, S. 166 f.]

Eine Emotion kann sich auf verschiedene Art und Weisen ausdrücken. Die Reaktionen, die auf eine Emotion folgen, werden in einem Komponentenmodell (Abb. 9) zusammengefasst. Diese Komponenten werden in vielen Ansätzen zur Unterscheidung einzelner Emotionen wieder aufgegriffen. Ebenfalls lassen sich mithilfe der fünf Komponenten die Emotionen von anderen Gefühlen abgrenzen. Im Folgenden wird auf die einzelnen Komponenten eingegangen. Zur Visualisierung kann

jederzeit die Abbildung 9 genutzt werden, welche das Geschriebene grob zusammenfasst. [20, S. 167 ff.]

Abb. 9: Das Komponentenmodell der Emotionen [20, S. 168]

Die erste Komponente ist die Erlebenkomponente. Diese sagt aus, dass Menschen die **Gefühle subjektiv erleben** und entsprechend auch subjektiv fühlen. So empfindet jeder Mensch zum Beispiel die Emotion Freude ganz unterschiedlich. Aufgrund dessen, dass diese Gefühle nicht objektiv betrachtet und untersucht werden können, wird diese Komponente von Emotionsforschern meistens eher als unwichtig oder unwissenschaftlich eingestuft. [20, S. 168 ff.]

Eine weitere Komponente ist die kognitive Komponente. Die kognitive Komponente beschreibt, dass das **Erleben von Gefühlen durch Bewerten und Urteilen geprägt** wird. Hierbei können verschiedene Emotionen entstehen, je nachdem, ob Ereignisse als positiv oder negativ eingestuft werden. Es muss jedoch nicht nur diese Einschätzung die eigenen Emotionen beeinflussen. Das Erleben von Emotionen kann mit mehreren kognitiven Vorgängen verknüpft sein. Hierbei kann ein

Erfolg in einer Klausur verschiedene Emotionen hervorrufen. Stolz kann durch die eigene gute Leistung hervorgerufen werden. Erleichterung kann hervorgerufen werden, wenn die Fragen günstig ausgewählt worden sind, und Dankbarkeit kann hervorgerufen werden, wenn der Sitznachbar Hilfestellung geleistet hat. Diese kognitiven Vorgänge müssen uns nicht zwingend bewusst sein. [20, S. 170]

Die physiologische Komponente zeigt, dass Emotionen unterschiedliche physiologische Vorgänge im Körper auslösen können. So sollen Emotionen Veränderungen von Aktivitäten im *autonomen Nervensystem* aufzeigen. Ein erhöhter Herzschlag und eine erhöhte Hautleitfähigkeit sollen beispielsweise von Angst ausgelöst werden. Jedoch haben Emotionsforscher herausgefunden, dass sich allein durch diese Faktoren nicht immer Emotionen voneinander unterscheiden lassen. Es lässt sich nur ein Unterschied zwischen positiven und negativen Emotionen erkennen. Neurobiologen fanden heraus, dass sich emotionale Zustände in den neurobiologischen Substraten wiederfinden, welche netzwerkartig Areale im Gehirn umspannen. Dies bedeutet, dass **eine Emotion oft mehrere Bereiche des Gehirns aktiviert** und nicht eine Zone im Gehirn für genau eine Emotion zuständig ist. [20, S. 170 ff.]

Die Ausdruckskomponente soll zeigen, dass sich Emotionen anhand von **Mimik, Gestik, Haltung und Stimme** erkennen lassen können. Ein besonders stark untersuchter Faktor ist die Mimik. Zwischen den verschiedenen Kulturen lässt sich kein Unterschied im mimischen Ausdruck erkennen. Die emotionsspezifische Mimik ist zudem mit hoher Wahrscheinlichkeit angeboren. Dies zeigte sich bei Kindern, die von der Geburt an blind sind. Jedoch lässt sich trotzdem kein Schluss vom mimischen Ausdruck auf die Befindlichkeit ziehen, da dies von dem sozialen Kontakt und den sozialen Darstellungsregeln abhängig ist. [20, S. 172 ff.]

Die letzte Komponente ist die motivationale Komponente, welche besagt, dass **Emotionen mit bestimmten Handlungen einhergehen**, die sich bewährt haben bei der Bewältigung von bestimmten Herausforderungen. Sie werden deshalb auch als Verhaltensstrategien für diese Herausforderungen gesehen. Bei den Verhaltensstrategien beziehen sich die Emotionen auf abstrakte Verhaltensfunktionen, wie zum Beispiel Vermeidung, aber nicht auf ein bestimmtes Verhalten. Deshalb ist

es situationsabhängig, wie eine Person reagiert. Dennoch haben die verschiedenen Verhaltenstaktiken dieselben oder zumindest ähnliche Funktionen. [20, S. 174 f.]

Es besteht die Annahme, dass Emotionen eine *Korrelation* zwischen den einzelnen Komponenten aufweisen. Jedoch konnten Studien schwache, mittlere oder sogar keinen Zusammenhang zwischen den Komponenten feststellen. Weshalb bis heute noch bestenfalls ein schwacher Zusammenhang zwischen den Komponenten angenommen wird. [20, S. 175]

Eine Studie¹ von 2019 zeigt ähnliche Ergebnisse. Diese besagt, dass Personen ihre Emotionen auf verschiedene Art und Weise kommunizieren. Denn je nach Situation, Kultur oder sogar bei denselben Personen und Situationen können die vorher erwähnten Komponenten komplett unterschiedlich aussehen. Genauso können zum Beispiel dieselben Gesichtsausdrücke mehr als nur eine Emotion ausdrücken. Allerdings bringen die Gesichtsausdrücke eine Reihe an Informationen und sind dennoch wichtig für soziale Kommunikation. [2] [21]

Doch was genau sind nun Emotionen und wie viele gibt es? Es existieren ungefähr 34.000 Emotionen. Für uns Menschen sind diese eher nicht greifbar. Aus diesem Grund wurden Modelle erstellt, die dies vereinfachen sollen. Diese kategorisieren und sortieren Emotionen. Eines der bekanntesten Modelle ist das **Rad der Emotionen** von Robert Plutchik², welches in Abbildung 10 zu erkennen ist. Seine Theorie fasst er in zehn *Postulaten* zusammen. Diese werden im Folgenden aufgelistet und kurz erläutert:

- **Tiere und Menschen:** Das *limbische System* eines Menschen ähnelt dem von anderen Lebewesen, weswegen diese dieselben grundlegenden Emotionen aufzeigen.

1 Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements

2 Robert Plutchik (1927-2006) war ein US-amerikanischer Psychologe. [34]

- **Evolutionäre Geschichte:** Emotionen sind Teil der Evolution und haben verschiedene Ausdrucksformen bei verschiedenen Lebewesen entwickelt.
- **Überlebensprobleme:** Emotionen helfen uns bei der Bewältigung wichtiger Überlebensprobleme der Umwelt.
- **Prototypische Muster:** Emotionen identifizieren sich durch bestimmt prototypische Muster oder Elemente.
- **Grundlegende Emotionen:** Die primären Emotionen sind: Groll, Erwartung, Abneigung, Angst, Freude, Vertrauen, Traurigkeit und Überraschung.
- **Kombinationen:** Das Kombinieren primärer Emotionen lässt neue entstehen, zum Beispiel ergeben Freude und Vertrauen Liebe.
- **Hypothetische Konstrukte:** Primäre Emotionen sind Konstrukte oder Ideen, die dabei helfen eine bestimmte Erfahrung zu beschreiben.
- **Gegensätze:** Jede primäre Emotion hat ihren polaren Gegensatz, zum Beispiel ist Traurigkeit der Gegensatz von Freude.
- **Ähnlichkeiten:** Der Ähnlichkeitsgrad gibt an, welche Emotionen miteinander verwandt und welche das komplette Gegenteil sind.
- **Intensität:** Der Intensitätsgrad gibt die Änderung der Emotionen in ihrer Intensität an und erzeugt vielfältige Emotionen, die wir fühlen können, zum Beispiel geht Vertrauen von Akzeptanz bis hin zu Bewunderung. [25] [26]

Robert Plutchik ging davon aus, dass **acht grundlegende Emotionen** existieren: Groll, Erwartung, Abneigung, Angst, Freude, Vertrauen, Traurigkeit und Überraschung. Diese sollen die Grundlage für alle anderen Emotionen sein. In der Abbildung 10 sind sie jeweils in unterschiedliche Farben unterteilt worden. Die Nuance der jeweiligen Farbe zeigt auch die Intensität der jeweiligen Emotion an. Die intensivsten Emotionen stehen in der Mitte des Kreises. Nach außen hingehend lässt die Intensität einer Emotion nach. Zum Beispiel kennzeichnet die Farbe Gelb die Emotion Freude. Innerhalb des Kreises ist das Gelb am stärksten und die zugehörige Emotion ist Ekstase. Im mittleren Kreis ist die Emotion Freude. Ganz außen ist die schwächste Nuance der Farbe Gelb und damit auch die geringste Intensität und die zugehörige Emotion ist Gelassenheit. Die primären Emotionen werden außerdem in die jeweiligen Gegenteile wie folgt gruppiert:

- Freude und Traurigkeit
- Vertrauen und Abneigung
- Angst und Groll
- Überraschung und Erwartung [26] [27]

Abb. 10: Plutchiks Rad der Emotionen [22]

Diese sind auch am Rad der Emotionen zu erkennen, da sich diese genau gegenüberliegen. Des Weiteren lassen sich die Kombinationen von den primären Emotionen in den Zwischenräumen der Emotionen in Abbildung 10 erkennen. Zum Beispiel ergibt die Kombination von Freude und Erwartung Optimismus. [26]

Nach Robert Plutchik werden **Emotionen durch bestimmte Reize aktiviert** und **lösen bestimmte Verhaltensmuster aus**. In Tabelle 1 ist zu erkennen, welche Emotion welches Verhalten auslöst, um eine bestimmte Funktion zu erreichen. Angst soll somit durch Bedrohung ausgelöst werden. Zum eigenen Schutz wird eine Person dann eine Situation oder ein Objekt meiden oder davor fliehen. [26]

Tabelle 1: Emotionen, ihre Funktionen und das dadurch ausgelöste Verhalten. Eigene Tabelle nach [20, S. 169 f.] und [141]

EREIGNIS		EMOTION		VERHALTENSIMPULS		FUNKTION
Potenzieller Partner	→	Freude	→	Werben, Paaren	→	Reproduktion
Gruppenmitglied	→	Vertrauen	→	Umsorgen, Teilen	→	Affiliation
Bedrohung	→	Angst	→	Fliehen, Vermeiden	→	Schutz
Unerwartetes Objekt	→	Überraschung	→	Stoppen, Alarmieren	→	Orientierung
Verlust einer Bezugsperson	→	Traurigkeit	→	Ruf nach Hilfe	→	Wiedervereinigung
Ungenießbares Objekt	→	Abneigung	→	Ausspucken, Wegstoßen	→	Zurückweisung
Hindernis, Blockade	→	Groll	→	Beißen, Schlagen	→	Zerstörung
Neues Territorium	→	Erwartung	→	Untersuchen, Erkunden	→	Exploration

Dies bedeutet, dass unsere Emotionen unbewusst Überlebensverhaltensweisen hervorrufen. Die nach Plutchik definierten Ereignisse und Verhaltensimpulse sind jedoch sehr unvollständig. Menschen empfinden Freude nicht nur bei einem Partner und der Fortpflanzung, sondern auch bei verschiedenen Objekten und Aktivitäten, welche uns subjektiv Freude bereiten können, wie zum Beispiel ein Geschenk oder ein Sonnenuntergang. Dies lässt sich nicht nur auf die Emotion Freude beziehen, sondern auch auf die anderen Emotionen. Somit zeigt dieser Ansatz deutliche Schwächen auf. [26]

Ein weiteres Modell ist das **Circumplex Modell** von James A. Russell¹ und Lisa Feldman Barrett². Dieses zweidimensionale Kreismodell (s. Abbildung 11) zeigt emotionale Zustände um zwei orthogonale Dimensionen herum, die sich kreisförmig anordnen. Die Dimensionen sind in dem Modell als *Valenz* (angenehm bis unangenehm) und *Erregung* (ruhig bis erregt). Die Erregung ist in der Abbildung als Aktivierung und Deaktivierung gekennzeichnet. Es wird davon ausgegangen, dass Emotionen durch unterschiedliche Kombinationen von Erregung und *Valenz* definiert werden. Im Inneren des Kreises finden sich somit grundlegende *affektive*

Abb. 11: Das Circumplexmodell nach Lisa Feldman Barrett und James A. Russell [20, S. 169]

Zustände und auf dem äußeren Kreis werden typische emotionale Zustände angeordnet. Überraschung ist beispielsweise ganz oben angeordnet. Damit hat diese

-
- 1 James A. Russell (geboren 1947) ist ein US-amerikanischer Psychologe, dessen Arbeit sich auf Emotionen fokussiert. [88]
 - 2 Lisa Feldman Barrett ist Professorin für Psychologie. [89]

Emotion eine neutrale *Valenz*, ist also weder angenehm noch unangenehm. Aber Überraschung hat eine hohe Erregung. Freude hingegen ist auf der angenehmen Seite lokalisiert, mit mittlerer Erregung. [20, S. 169 f.]

Eventuell lässt sich bei diesem Modell auch eine dritte Dimension erkennen. Die gestrichelten orthogonalen Linien in der Mitte des Kreises sollen eine dritte Dimension, bestehend aus einem niedrigen bis hohen positiven *Affekt* (PA) und einem niedrigen bis hohen negativen *Affekt* (NA), darstellen. Bei denen ist man sich allerdings nicht sicher, ob diese im direkten Zusammenhang mit der *Valenz* stehen. Damit wäre die dritte Dimension irrelevant. [20, S. 169 f.]

Die beiden vorgestellten Modelle unterscheiden sich in wesentlichen Punkten. Das Modell von Plutchik beschreibt die Emotionen untereinander, sodass man diese miteinander vergleichen kann. Des Weiteren können in diesem Modell neue Emotionen aus den primären Emotionen entstehen. Somit besteht ein Zusammenhang zwischen den Emotionen. In dem Modell von Barrett und Russell wird eine Wertung für jede einzelne Emotion vorgenommen. Je nach Wertung werden die Emotionen angeordnet und können nur aufgrund dessen miteinander verglichen werden. Dieses Modell zeigt jedoch keinerlei offensichtliche Zusammenhänge zwischen den Emotionen.

Dieses Kapitel zeigt, was Emotionen sind, wodurch sich diese definieren und wie diese voneinander abgegrenzt werden. Die gewonnenen Informationen sollen als Grundlage für das Verständnis der Algorithmen in Kapitel 5.3 fungieren. Im folgenden Kapitel wird die Kommunikation zwischen Mensch und Roboter erörtert. Hierbei wird unter anderem untersucht, welche Rolle Emotionen bei humanoiden Robotern spielen und wie humanoide Roboter aktuell auf Menschen wirken.

3.2. Kommunikation zwischen Mensch und Roboter

Die Kommunikation zwischen Mensch und Maschinen spielt schon heute eine große Rolle in unserem alltäglichen Leben. Dabei fragen wir Siri¹ nach dem Weg nach Hause, Alexa² bitten wir das Licht auszuschalten oder sagen dem Google Home Gerät³ es solle Musik abspielen. Das ist die übliche Form, in der wir momentan mit Geräten in unserer näheren Umgebung, zum Beispiel zu Hause, kommunizieren. Auch eine weitere Form der Kommunikation ist uns bereits bekannt, nämlich die Kommunikation über Texteingabe. Sogenannte *Chatbots* sollen die Informationsbeschaffung auf Webseiten, wie Onlineshops, vereinfachen. Doch meistens erkennt man auch schnell Schwachstellen in der Kommunikation mit Maschinen. Entweder wird das Gesagte nicht richtig erkannt, die Internetverbindung ist nicht vorhanden, die abzurufende Information ist nicht verfügbar, oder ein Gespräch kann nicht fortgeführt werden. Viele dieser Schwachstellen werden über die Jahre hinweg stetig besser.

Die Kommunikation zwischen einem Menschen und einem Roboter ist jedoch oft eine andere als mit einem einfachen Gerät. Hierbei geht es neben der Kommunikation auch um die **Interaktion**. Doch wie werden Kommunikations- und Interaktionsarten vom Menschen wahrgenommen und welche Folgen können diese haben? Wichtig dabei ist es, dass sowohl die Kommunikation als auch die Interaktion zu den Aufgaben des Roboters passt. Ein Roboter in einer Produktionsfabrik wird kaum oder gar keine Interaktion mit Menschen haben, sodass dessen Kommunikation zum Menschen sehr sporadisch ausfällt. Ein Roboter in der Therapie hingegen muss eine menschnahe Kommunikation und Interaktion beherrschen. Die Therapie-Robbe Paro⁴ kann demenzkranken Personen helfen, sich anderen zu öffnen.

1 Siri ist der Sprachassistent von Apple (<https://www.apple.com/>).

2 Alexa ist der Sprachassistent von Amazon (<https://www.amazon.com/>).

3 Google Home Geräte sind intelligente Lautsprecher inklusive eines Sprachassistenten der Firma Google (<https://www.google.com/>).

4 Therapie-Robbe Paro <http://www.parorobots.com/>

Paro ahmt je nach Berührung und Gesagtem eine echte Robbe mit Geräuschen und Bewegungen nach. Somit hat die Robbe einen beruhigenden Einfluss auf die Personen. Paro ist in Abbildung 12 zu erkennen. [1] [39] [41]

Abb. 12: Therapie Robbe Paro [59]

Forscher und Wissenschaftler finden verschiedene Ansätze zur Kommunikation mit Robotern. Tamim Asfour, ein Professor des Karlsruher Institut für Technologie, sagt: "Wenn ich weiß, dass der Roboter sich wie ein Mensch bewegt, dann kann ich als Interaktionspartner mich besser darauf einstellen.". Dies gilt für Bewegungen genauso wie für die Art der Kommunikation, Gestik oder Mimik. Das bedeutet auch, dass es uns leichter fällt, einen Roboter zu verstehen und ihn zu bedienen je ähnlicher er dem Menschen ist. Auf der anderen Seite jedoch existiert das Phänomen "**Uncanny Valley**". Dieses Phänomen ist seit den Siebzigerjahren bekannt und sagt aus, dass Roboter sympathischer wirken, je menschenähnlicher sie sind. Ab einem bestimmten Punkt jedoch wirken sie meist unheimlich. Das unheimliche Gefühl verfliegt erst ab dem Punkt, an dem ein Roboter nicht mehr von einem Menschen zu unterscheiden ist.

Forscher stellten außerdem fest, dass ein Roboter am **unheimlichsten** empfunden wird, wenn er **eigene Gefühle und Bedürfnisse** hat. Roboter in der Pflege lösen dieses Empfinden jedoch nicht bei Personen aus. Für einen humanoiden Roboter muss also, je nach Einsatzgebiet, ein akzeptabler Grad der Menschlichkeit gewählt werden. Hierfür sollen sich Roboter mit stark vereinfachten Gesichtszügen, die klar

als Maschine erkennbar sind, am besten eignen. Denn wenn solche Roboter menschliche "Fehler" begehen wirken sie sympathischer, als wenn sie dem Menschen so ähnlich sehen, dass man sie kaum noch unterscheiden kann. Als menschliche "Fehler" können zum Beispiel längere Reaktionszeiten oder auch falsche Reaktionen gesehen werden. [1] [43]

Abb. 13: Das Trolley-Problem [45]

Eine Studie zeigt, wie viel Empathie Menschen für einen humanoiden Roboter empfinden können. Diese Studie basiert auf dem **Trolley-Problem** (Abb. 13), einem moralischen Gedankenexperiment von Karl Engisch¹ (neuere Fassung von Philippa Foot²). Bei diesem Problem fährt ein Zug auf den Schienen und kann nicht angehalten werden. Der Zug kann nun auf zwei verschiedene Gleise geleitet werden. Auf dem einen befinden sich mehrere Personen, die überrollt werden würden, und auf dem anderen befindet sich eine Person bzw. weniger Personen als auf dem anderen Gleis. Nun muss eine Person entscheiden, auf welchem Gleis der Zug fahren soll. In der Studie wurde dieses Dilemma etwas abgewandelt, indem die einzelne Person auf den Schienen mit einem Roboter ausgetauscht worden ist. Je menschlicher dieser Roboter wirkte, desto eher waren die Personen bereit, die Gruppe von Personen zu opfern und den humanoiden Roboter zu retten. [1] [47] [48]

1 Karl Engisch (1899-1990) war ein deutscher Strafrechtler. [50]

2 Philippa Foot (1920-2010) war eine britische Philosophin. [49]

Dass Roboter Emotionen erkennen, entsprechend verarbeiten und auch selbst imitieren können, wird als **Affective Computing** bezeichnet. Affective Computing ist ein Teil künstlicher Intelligenz. Sophia (Abb. 14), ein menschenähnlicher Roboter, zeigt schon heute, wie genau dieser Teil der künstlichen Intelligenz aussehen kann. Sophia kann, neben vielen anderen intelligenten

Abb. 14: Roboter Sophia [19]

Handlungen, Emotionen deuten und über 60 menschliche Mimiken imitieren. Sophia selbst sagt: "Soft Skills sind wichtig, um die richtige Entscheidung zum richtigen Zeitpunkt zu treffen.". [51]

Dieses Kapitel zeigt, dass viele Roboter nicht mehr nur als Maschine eingesetzt werden, sondern auch immer mehr als direkter Kommunikationspartner für den Menschen dienen. Hierfür ist es wichtig, dass Roboter eine **gewisse Menschlichkeit** mit sich bringen. Die Emotionserkennung kann in dieser Hinsicht also sehr vorteilhaft sein und Roboter menschlicher in ihrer Interaktion mit den Menschen wirken lassen. Dabei muss er jedoch keinesfalls perfekt sein. Ein Zitat fasste dies sehr gut zusammen: "[...] Und seien wir mal ehrlich: auch uns Menschen fällt es oft schwer, zu verstehen, was unsere Freunde und Familien empfinden – wie sollen Roboter diese Aufgabe dann bewältigen? Das menschliche Verhalten ist vielfältig, wechselhaft und oft unberechenbar.[...]" [52].

Ein Roboter wirkt eben auch menschlich, wenn er (unbeabsichtigt) Fehler begeht. Und doch sollten Roboter Emotionen menschenähnlich erkennen können. Könnte ein Roboter zum Beispiel nur vier Emotionen erkennen, kann dies eventuell auch einen Einfluss auf die Interaktion zwischen uns Menschen haben. Die Kategorisierung der Emotionen gestaltet sich jedoch generell schwierig. Für Emotionen gibt es keine einheitliche Definition. Sie sind komplex und es ist noch nicht komplett und detailliert erforscht, wie sie gemessen werden können und ob sie von Mensch zu Mensch universell sind oder zum Beispiel kulturelle Unterschiede haben. [53] [54]

In diesem Kapitel wurde erkannt, dass Emotionen komplex sind, sich von Mensch zu Mensch unterscheiden und auf verschiedene Art und Weise definiert werden können. Jedoch sind Emotionen tief in unserem Unterbewusstsein verankert und beeinflussen somit auch unsere Reaktionen auf Situationen. Dieses Erkenntnis kann genutzt werden, damit Roboter Menschen besser verstehen und auch entsprechend auf Situationen reagieren können. Somit kann die gesamte Kommunikation zwischen Mensch und Roboter verbessert werden.

In den folgenden Kapiteln soll die Kategorisierung der Emotionen und die Unterscheidung derer beachtet werden. Des Weiteren soll darauf geachtet werden, dass ein Roboter durch die Emotionserkennung nicht zu menschlich wirkt.

4. Technische Grundlagen

In diesem Kapitel werden die technischen Grundlagen erläutert. Hierzu wird zuerst ROS-E vorgestellt. Anschließend werden Lizenzen von *Open-Source*-Software erläutert. Zuletzt werden benötigte Begrifflichkeiten aus Teilbereichen der künstlichen Intelligenz erklärt.

4.1.ROS-E

ROS-E ist prototypisch im Rahmen einer Bachelor-Thesis¹ an der Technischen Hochschule in Wildau entstanden. Weiterentwickelt wurde er im Laufe der Zeit von Praktikanten und Werkstudenten des RoboticLabs der TH Wildau und ist damit eine reine Studentenentwicklung. Er wurde als ein **Tischroboter** zum Einsatz in **Assisted Living** konzipiert. ROS-E wurde entwickelt, um neben den an der TH Wildau genutzten Robotern Nibo2² und SoftBank Robotics NAO³ und Pepper⁴, eine weitere Alternative zu bieten. ROS-E soll all die positiven Komponenten der Roboter zusammenfassen und die negativen bewusst weglassen. Des Weiteren besteht die Möglichkeit bei einem selbst entwickelten Roboter die Soft- und Hardware beliebig anzupassen oder zu erweitern. Dies kann je nach Anwendungsszenario in Praktika oder für wissenschaftliche Arbeiten angepasst werden. Ebenfalls können Dinge, die sich im Nachhinein als nicht praktikabel herausstellen noch geändert werden. [40]

-
- 1 Valentin Schröter: Prototypischer Bau eines Tischroboters als Plattform für zukünftige Entwicklungs- und Forschungsprojekte im Bereich des Ambient Assisted Livings und als Alternative zum NAO-Roboter; Bachelorarbeit; Studiengang Telematik
 - 2 Nibo2 <http://www.nicai-systems.com/de/nibo2>
 - 3 NAO <https://us.softbankrobotics.com/nao>
 - 4 Pepper <https://us.softbankrobotics.com/pepper>

Der humanoide Roboter besteht aus einem Gehäuse, in welchem die zentrale Recheneinheit verbaut ist. Der Kopf wurde mit **2 Servomotoren** am ROS-E befestigt, um diesen bewegen zu können. Zu den Aktoren von ROS-E gehören ein **Display**, **Augen** und ein **Lautsprecher**. Das Touch Display dient zum Darstellen von Informationen und als Kommunikationsmittel zwischen Mensch und Roboter. Die Augen sind mit zwei 1,3 Zoll OLED Displays realisiert. Dies hat den Vorteil, dass die Augen verschiedene Gesichtsausdrücke darstellen können, wie auch in Abbildung 15

Abb. 15: ROS-E mit neutralen Augen [139]

und in Abbildung 16 zu sehen. Zu den Sensoren von ROS-E gehören eine **Kamera**, eine **Geräuschwahrnehmung** und **Berührungssensoren**. Einer dieser Berührungssensoren übernimmt unter anderem die Funktion des Ein- und Auschalters. [40]

ROS-E verfügt über einen **LAN-Anschluss** und über ein **WLAN-Modul** und damit die Möglichkeit Internetzugang zu nutzen. Ebenfalls besitzt der Roboter ein **Bluetooth-Modul**. Aufgrund der zentralen Recheneinheit ist auf dem ROS-E ein Ubuntu Betriebssystem installiert. Nvidia stellt hier jedoch ein spezielles Image bereit, welches weitere Softwarepakete und Frameworks beinhaltet. Zu diesen gehören unter anderem JetPack, TensorFlow und OpenCV. Die installierte **Ubuntu**-Version ist 18.04. Des Weiteren wurde das Framework ROS (Robot Operating System) auf dem ROS-E installiert. ROS ist ein Framework, welches direkt für Roboter entwickelt worden ist. Dieses soll es vereinfachen verschiedene Programmkomponenten

miteinander zu verknüpfen, aber auch das Zusammenspiel von Software und Hardware zu vereinfachen. [40]

ROS-E soll als Kumpel vor allem bei älteren Personen, Familien und Kindern eingesetzt werden, da er die Möglichkeiten einer natürlichen Kommunikation bietet. Die Kommunikation kann sowohl über sprachliche Wege führen, als auch über das dazugehörige Touch Display. Einsatzszenarien für ROS-E sind aufgrund dessen sehr vielfältig. Innerhalb von Altersheimen könnte er an bestimmte Termine, Aufgaben oder an die Einnahme von Medikamenten erinnern. Des Weiteren ist denkbar, dass mit ROS-E Spiele und Sportprogramme durchgeführt werden können. Aber auch das Anhören von Musik oder das Anschauen von Videos können als Unterhaltung über ROS-E dienen. Ebenso kann er als Kommunikationsmittel fungieren, indem mithilfe von ihm Videoanrufe mit Freunden, der Familie oder auch mit einem Arzt geführt werden können. Des Weiteren könnte ROS-E allerhand Informationen liefern. Zu diesen Informationen können das Wetter, anstehende Ereignisse, der aktuelle Essensplan oder auch interaktive Rezepte gehören. [40]

Abb. 16: ROS-E mit einem anderen Ausdruck in den Augen [139]

4.2. Lizenzen

In diesem Abschnitt werden verschiedene Lizenzen genannt und genauer erläutert, in Vorbereitung auf die verschiedenen *Open-Source*-Programme. Somit wird klar, inwiefern diese genutzt, verändert und veröffentlicht werden können.

MIT-Lizenz

Die MIT-Lizenz erteilt jeder Person, die eine Kopie dieser Software und der Dokumentation erhält, die Erlaubnis, die Software uneingeschränkt zu nutzen, zu kopieren, zu verändern, zusammenzufügen, zu veröffentlichen, zu vertreiben, zu unterlizenzieren oder Kopien zu verkaufen. Die Software wird ohne jegliche Gewährleistung zur Verfügung gestellt. Die Autoren und *Copyright*-Inhaber haften für keine Schäden oder andere Haftungen, die in jeglicher Art und Weise in den Zusammenhang mit der Software gebracht werden kann. Der *Copyright*-Hinweis muss in allen Kopien oder in den wesentlichen Teilen der Software enthalten sein. Dieser Hinweis ist vollständigerweise in Anhang B einzusehen. [76]

Apache-2.0 Lizenz

Die Apache-2.0 Lizenz ist eine von "The Apache Software Foundation"¹ im Jahre 2004 bereitgestellte Lizenz für *Open-Source*-Software. Software, die mit dieser Lizenz lizenziert ist, kann uneingeschränkt genutzt, kopiert, verändert, zusammengefügt, vertrieben oder verkauft werden. In der Software muss eindeutig auf die Lizenz hingewiesen werden. Die Lizenzgeber und alle Mitwirkenden stellen die Software ohne jegliche Gewährleistung zur Verfügung. Des Weiteren haftet kein Lizenzgeber oder anderer Mitwirkender für jegliche Schäden oder andere Haftungen. Dennoch kann Support, Garantie, Entschädigung oder eine Haftung für diese Lizenz angeboten werden, auch gegen Gebühr. Jedoch muss jeder andere Mitwirkende von der Haftung ausgeschlossen sein. Die Apache-Lizenz sollte in der Arbeit in einer Datei "LICENSE" aufgeführt sein. Für bestimmte Dateien kann das

1 The Apache Software Foundation (ASP): Anbieter für *Open-Source*-Software. <http://www.apache.org>

Copyright in Anhang C verwendet werden. Dafür müssen die Daten in eckigen Klammern mit den eigenen ausgetauscht werden. [78]

AGPL-3.0 Lizenz

Die AGPL-3.0 (Affero General Public License, Version 3) ist eine Softwarelizenz der Firma GNU¹ aus dem Jahr 2007 mit Copyleft. Copyleft ist eine Methode, die dazu beitragen soll, dass alle Modifikationen oder Erweiterungen eines freien Programms ebenfalls frei sind. Ein freies Programm wird als frei bezeichnet, wenn das Programm beliebig ausgeführt, verbessert, veröffentlicht, die Funktionsweise untersucht und die Art der Datenverarbeitung angepasst werden kann. Damit grenzt sie sich von anderen Lizenzen ab. Ebenso muss der Quellcode eines freien Programms zugänglich sein oder auf Wunsch geteilt werden. Für die lizenzierten Programme besteht keinerlei Gewährleistung. Kein Urheberrechtinhaber ist für Schäden haftbar. In Anhang D sind die Vermerke ersichtlich, die in den Quellcode eingebunden werden müssen. GNU empfiehlt diese an den Anfang jeder Datei zu schreiben, zumindest sollte jedoch die erste Zeile (*Copyright*) jeder Datei hinzugefügt werden. [79] [80] [81]

1 GNU ist ein linuxbasiertes Betriebssystem. <https://www.gnu.org/>

4.3. Künstliche Intelligenz

"Künstliche Intelligenz (KI) beschäftigt sich mit Methoden, die es einem Computer ermöglichen, solche Aufgaben zu lösen, die, wenn sie vom Menschen gelöst werden, Intelligenz erfordern." [142]

Die KI kann in viele verschiedene Teilgebiete unterteilt werden. Im Folgenden werden bestimmte Teilgebiete und Begrifflichkeiten in der KI erläutert. Diese werden in Kapitel 5 verwendet.

4.3.1. Machine Learning

Machine Learning (Maschinelles Lernen) ist ein Teil der KI und bezeichnet die Anwendung von Verfahren, durch die Systeme die Fähigkeit erlangen selbstständig anhand von Eingabewerten das Lösen von Problemen kontinuierlich zu verbessern. [91] [92]

4.3.2. Deep Learning

Deep Learning ist ein Teil der KI und Machine Learning. Deep Learning arbeitet mit neuronalen Netzen mit großen, unstrukturierten Datensätzen. [109]

4.3.3. Künstliches Neuronales Netz (KNN)

Ein KNN ist ein künstliches neuronales Netz (Abb. 17). Der Aufbau eines KNNs gleicht dem eines **menschlichen Gehirns**. Es besteht aus Neuronen, die Informationen aufnehmen, modifizieren, weiterleiten oder als **Ergebnis** ausgeben. Neuronen bestehen aus **gewichteten Eingängen**, einer **Aktivierungsfunktion**, welche die Eingabe beeinflusst, und einer **Ausgabe**. Die Informationen können von außen oder von anderen Neuronen geleitet werden. Die Informationen werden über Kanten gesendet. Kanten verbinden zwei Neuronen miteinander. Jede Kante hat eine bestimmte Gewichtung, die aussagt, wie viel Einfluss ein Neuron auf ein Anderes haben kann. Die Gewichtungen können positiv, negativ oder null sein. Bei einer positiven Gewichtung hat das Neuron eine erregende Wirkung. Bei einer negativen Gewichtung hat das Neuron demzufolge eine hemmende Wirkung. Ist die Gewichtung null, hat das Neuron keinerlei Wirkung auf ein anderes. [82] [105]

Abb. 17: Künstliches Neuronales Netz [143]

Ein KNN besteht aus **drei** miteinander verbundenen **Schichten**: Eingabe-Schicht, verdeckte Schicht/en, Ausgabe-Schicht. Die Eingabe-Schicht sendet Informationen der Eingabe-Neuronen an die verdeckte Schicht. Die verdeckte Schicht kann aus mehreren Schichten bestehen und leitet Informationen an die Ausgabe-Schicht. [82]

Ein wichtiger Parameter eines KNNs ist die **Lernrate**. Die Lernrate ist ein festgelegter Wert, der angibt wie die Gewichtungen der Kanten während des Trainings verändert werden. Eine geringe Lernrate erfordert mehrere Durchläufe der Datensätze während des Trainings, da die Gewichtungen nur minimal geändert werden. Eine hohe Lernrate erfordert weniger Durchläufe der Datensätze während des Trainings, aber die Gewichtungen werden in größeren Schritten geändert. [82] [132] [133]

4.3.4. Convolutional Neural Network (CNN)

Ein CNN ist ein KNN, welches für Machine Learning und die **Verarbeitung von Bild- und Audiodaten** vorgesehen ist. Die Verarbeitung von Bild- und Audiodateien ist mit einem klassischen KNN sehr rechen- und zeitintensiv. Aus diesem Grund wurde ein CNN entwickelt. Der Aufbau ist dem einer Sehrinde eines Gehirns nachempfunden. Dieser Aufbau ist in Abbildung 18 dargestellt. Das CNN besteht aus den folgenden Schichten:

- **Convolutional-Schicht**
- **Pooling-Schicht**
- **vollständig vermaschte Schicht**

Auf jede Convolutional-Schicht muss eine Pooling-Schicht folgen. Diese können auch mehrmals vorkommen. Die Convolutional- und Pooling-Schicht sind beide lokal vermaschte Teilnetze. Diese beiden Schichten

können beliebig oft aufeinander folgen. Zum Schluss muss eine vollständig vermaschte Schicht (Fully-Connected Layer) folgen. [72] [130]

Abb.18: Aufbau eines CNNs [131]

Die Convolutional-Schicht, auch Faltungsschicht genannt, erkennt und extrahiert Merkmale der Eingabedaten. Die Eingabedaten werden während der Verarbeitung in die Form von Matrizen gebracht. Die Matrizen haben die Form: Breite x Höhe x Kanäle. [72] [130]

Die Pooling-Schicht eliminiert nicht-relevante Informationen indem sie Merkmale verdichtet und die Auflösung der Merkmale reduziert. Durch diese Schicht wird die Datenmenge reduziert und die Berechnung wird erhöht. [130]

Die vollständig vermaschte Schicht verbindet jedes Merkmal der vorherigen Schichten mit jedem Ausgabemerkmal. Die Menge der Neuronen richtet sich danach, wie viele Klassen oder Objekte unterschieden werden sollen. Die Neuronen können auf mehreren Ebenen angeordnet sein. [130]

4.3.5. Mehrlagiges Perzeptron (MLP)

Das MLP ist ein Modell eines KNNs, um Objekte zu klassifizieren. Ein Perzeptron besteht aus einem Neuron, das mehrere Eingänge und einen Ausgang besitzt. Die Eingänge werden gewichtet. Diese **Gewichtungen können** durch den Lernprozess **verändert werden**. Wenn ein gesetzter Schwellwert die Gewichtung aller Eingänge über- oder unterschreitet, wird der Zustand des Ausgangs verändert. Mehrmalige Perzeptren bestehen mehrere Schichten mit Neuronen, die untereinander vernetzt sind. Dies führt dazu, dass komplizierte Trennungen durchgeführt werden können. Auch dieses Modell muss mit bereits klassifizierten Daten angelernt werden. Damit kann es dann neuen Gesichtsmarkierungen Klassen zuordnen. [72] [125]

4.3.6. Recurrent Neural Network (RNN)

Ein Recurrent Neural Network (RNN) ist ein neuronales Netz, welches einen internen Speicher besitzt. Üblicherweise verarbeitet in einem neuronalen Netz jedes Neuron eine Aufgabe, welche unabhängig von anderen Informationen verarbeitet werden. Der Vorteil des RNNs ist, dass es während des Vorgangs **auf Informationen** der vorherigen Schritte **zugreifen** und diese bei Auswertung nutzen kann. RNNs werden unter anderem bei den Sprachassistenten von Apple und Google genutzt. [102]

4.3.7. Long Short-Term Memory (LSTM)

Ein Long Short-Term Memory (LSTM) ist eine Erweiterung des RNNs. Ein LSTM **verbessert den Speicher eines RNNs**. In einem LSTM werden drei Gates unterschieden: Input, Forget, Output. Das Input-Gate entscheidet, ob gewonnene Informationen im Eingang genutzt werden sollen. Das Forget-Gate entscheidet, ob die gewonnenen Informationen gelöscht werden können, falls sie nicht wichtig sind. Das Output-Gate entscheidet, ob die Informationen den Ausgang in dem aktuellen Schritt beeinflussen. Damit verbessert ein LSTM die Genauigkeit bei einem kürzeren Training. [102]

4.3.8. Deep Neural Network (DNN)

Ein Deep Neural Network (DNN) bezeichnet ein KNN, welches **viele verdeckte Schichten** besitzt. Damit kann es komplexere Aufgaben lösen. [82]

5. Stand der Technik

In diesem Kapitel werden die Möglichkeiten für einen hybriden Algorithmus zur Emotionserkennung aufgezeigt. Hierzu werden zuerst die Eingabeströme bestimmt, mit denen Emotionen erkannt werden können. Daraus werden die Grundpfeiler der Emotionserkennung erstellt. Diese beinhalten die am häufigsten genutzten Eingabeströme, aber nur die, die für die Anwendungsszenarien mit einem humanoiden Roboter geeignet sind. Im nächsten Schritt werden die verschiedenen Lizenzen erörtert, um einen Eindruck davon zu bekommen, wie die *Open-Source*-Projekte verwendet werden können. Des Weiteren werden ausgewählte *Open-Source*-Projekte und Dienste vorgestellt. Außerdem werden verschiedene Algorithmen für die drei Grundpfeiler der Emotionserkennung erläutert. Im letzten Schritt wird erläutert, welche Techniken oder Algorithmen große Firmen für ihre Emotionserkennung nutzen.

5.1. Die drei Grundpfeiler der Emotionserkennung

Für die Emotionserkennung existieren bereits die verschiedensten Dienste, *Open-Source*-Projekte und Algorithmen. Diese basieren sowohl auf verschiedenen Eingabeströmen für die Erkennung von Emotionen als auch auf verschiedenen Ansätzen für die Verarbeitung dieser. Meist wird auf Grundlage von **Bildverarbeitung** eine Emotionserkennung implementiert. Doch auch **Audio** oder **Text** werden gern als Eingabestrom zur Emotionserkennung genutzt, je nach Anwendung dieser. Eher selten werden **Eye-Tracking**, **EDA**, **EEG**, **EKG** oder **EMG** als Eingabeströme genutzt. Dies liegt wahrscheinlich daran, dass diese weniger praktikabel sind und somit nicht in alltägliche Produkte eingebunden werden. Des Weiteren kann kein natürliches Gespräch zwischen Mensch und Maschine entstehen, sobald die Person an zusätzliche Sensoren angeschlossen werden muss. Auch die Machbarkeit im Rahmen dieser Thesis stehen zusätzliche Sensoren außer Frage. [27] [65] [66] [67] [68] [69] [70] [71] [72]

Ein weiterer Faktor in der Emotionserkennung kann das **Alter** einer Person sein. Eine wissenschaftliche Arbeit zeigte in der sprachbasierten Emotionserkennung eine Verbesserung von 7 % durch die Einbeziehung des Alters einer Person. Somit erreichte diese Emotionserkennung eine Genauigkeit von 96 %. Des Weiteren kann auch das Einbeziehen eines **Geschlechts** zu einer verbesserten Emotionserkennung beitragen. Eine andere wissenschaftliche Arbeit zeigte ebenfalls in der sprachbasierten Emotionserkennung eine Verbesserung von bis zu 10 % bei bestimmten Emotionen. Lediglich bei der Emotion Traurigkeit wurde eine Verschlechterung erkannt. Diese beiden Faktoren können somit in die Emotionserkennung mit eingebunden werden, stellen aber keinen alleinigen Faktor für die Emotionserkennung dar. [74] [75]

Abb. 19: Die drei Grundpfeiler der Emotionserkennung

Aus diesen Gründen kristallisieren sich als häufigste Eingabeströme Bild, Ton und Text heraus. Diese drei Eingabeströme bilden somit die **Grundpfeiler der Emotionserkennung**. Das Alter und das Geschlecht bilden die schmalere Säulen in der Emotionserkennung. Durch diese Eingabeströme allein kann keine Emotions-

erkennung durchgeführt werden. Sie können jedoch zusätzliche Informationen liefern. In Abbildung 19 ist dies bildlich dargestellt.

5.2. Open-Source-Software, -Projekte und Dienste

In diese Kapitel werden verschiedene Dienste, *Open-Source*-Projekte und Algorithmen vorgestellt, die aus einzelnen oder auch aus mehreren der bereits genannten Säulen bestehen. Die Bedingungen für all diese sind, dass sie im Rahmen dieser Arbeit genutzt werden und gegebenenfalls prototypisch implementiert werden können.

5.2.1. Bild

Die erste Säule ist der Eingabestrom Bild. Diese soll mithilfe einer Kamera, Bildern oder Videos Emotionen bei Menschen erkennen können. Im Folgenden werden ausgewählte Dienste und Projekte vorgestellt.

Microsoft Azure Cognitive Services

Microsoft Azure¹ stellt einen KI-Dienst bereit, welcher Gesichtserkennung über Bildmaterial für verschiedene Einsatzszenarien bereitstellt. Unter anderem kann dieser Dienst Emotionen, das Alter und das Geschlecht einer Person erkennen. Die Informationen werden hierbei durch den Gesichtsausdruck interpretiert. Die zu erkennenden Emotionen sind: Wut, Verachtung, Ekel, Angst, Freude, Trauer, Überraschung und ein neutraler Gesichtsausdruck. Die Emotionen werden in Form von Dezimalzahlen angegeben. Somit kann der neutrale Gesichtsausdruck mit 0,2 und die Emotion Überraschung mit 0,8 in einem Bild bewertet sein. Es können

1 Microsoft Azure bietet cloudbasierte Dienste an. <https://azure.microsoft.com>

mehrere Personen in einem Bild erkannt werden. Die Daten sind in der Reihenfolge vom größten zum kleinsten Gesicht sortiert. Es wird davon ausgegangen, dass das größte Gesicht dem Betrachter am nächsten ist. Die Daten können über eine *POST-Anfrage* abgerufen werden und werden im JSON-Format zurückgegeben. Die Dokumentation der API¹ ist sehr ausführlich und mit Beispielen belegt. Eine kostenlose Demo² steht auf der Webseite von Microsoft Azure zur Verfügung. Mit der kostenfreien Version können maximal 30.000 bzw. 20 Anfragen in der Minute ausgeführt werden. [28] [83] [87]

FaceApi.js

FaceApi.js ist ein auf *JavaScript* basierendes *Open-Source*-Projekt zur Erkennung von Gesichtern, Anhaltspunkten im Gesicht, Emotionen und auch zur Erkennung des Alters und Geschlechts einer Person. Ebenso wird die Wahrscheinlichkeit der jeweiligen Emotion in Form von Dezimalzahlen angegeben. Somit ist es möglich, dass ein Gesichtsausdruck zu 70 % Freude ausdrückt und zu 30 % Überraschung. Die Erkennung der Emotionen begrenzt sich auf die Folgenden: Überraschung, Ekel, Angst, Trauer, Wut, Freude. Das Projekt ist lizenziert über eine MIT-Lizenz. Das Projekt wurde implementiert in Node.js³ und TensorFlow.js⁴. Die letzte Version ist vom 22. März 2020. Das Projekt kann in *JavaScript* oder *Typescript* genutzt werden. Die Dokumentation ist ausführlich und beinhaltet sowohl die Installation und Einrichtung als auch die eigentliche Nutzung des Projekts. Eben-

-
- 1 Dokumentation der API für Microsoft Azure Cognitive Services Gesichtserkennung: <https://westus.dev.cognitive.microsoft.com/docs/services>
 - 2 Microsoft Azure Cognitive Services Gesichtserkennung Demo: <https://azure.microsoft.com/de-de/services/cognitive-services/face/#demo>
 - 3 Node.js ist eine asynchrone, Event-basierte Laufzeitumgebung, die für die Entwicklung von Netzwerkanwendungen entworfen wurde. [90] (<https://nodejs.org>)
 - 4 TensorFlow ist eine *Open-Source*-Bibliothek für *maschinelles Lernen*. (<https://www.tensorflow.org>)

falls sind mehrere Beispiele aufgeführt. Auf vielen Blogs sind weitere Beispiele und Tipps zu dem Projekt gegeben. Die Community hingegen ist nur auf GitHub¹ aktiv und es wird nur ca. die Hälfte aller Fragen beantwortet. Getestet werden kann die Software mit der Demo². [85]

Stealing Ur Feelings

Stealing Ur Feelings ist durch einen 50.000\$ Preis von NetGain entstanden. Dies ist eine Zusammenarbeit zwischen Mozilla³, MacArthur Foundation⁴, Ford Foundation⁵, Knight Foundation⁶ und der Open Society Foundations⁷. Das Projekt hat viel Erwähnung gefunden, wie zum Beispiel bei Forbes⁸ oder bei CBC News⁹. Das Projekt kann Emotionen von Personen erkennen und angeben, zu wie viel Prozent diese erkannt worden sind. Auch hier werden diese Zahlen als Dezimalzahl ausgegeben. Zusätzlich kann die Software auch den IQ und das Einkommen einer Person bestimmen. Die

-
- 1 GitHub bietet eine Versionsverwaltung für Software-Projekte. (<https://github.com/>)
 - 2 FaceApi.js Demo: https://justadudewhohacks.github.io/face-api.js/face_and_landmark_detection
 - 3 Mozilla entwickelt Software-Produkte. (<https://www.mozilla.org>)
 - 4 Die MacArthur Foundation investiert in eine gerechtere, grünere und friedlichere Welt. [94] (<https://www.macfound.org/>)
 - 5 Die Ford Foundation investiert in die Verringerung von Armut und Ungerechtigkeit, demokratische Werte, internationale Zusammenarbeit und menschliche Leistungen. [95] (<https://www.fordfoundation.org/>)
 - 6 Die Knight Foundation investiert in den Journalismus und in die Demokratie. [96] (<https://knightfoundation.org/>)
 - 7 Die Open Society Foundations investieren in die Demokratie. [97] (<https://www.opensocietyfoundations.org/>)
 - 8 Forbes ist eine US-amerikanische Zeitschrift. [98] (<https://www.forbes.com/>)
 - 9 Die CBC (Canadian Broadcasting Corporation) News produzieren Nachrichten in Kanada. [99] (<https://www.cbc.ca/>)
 - 10 Stealing Ur Feelings Demo: <https://stealingurfeelin.gs/>

Software kann in einer sehr auffälligen Art und Weise in einer Demo¹⁰ getestet werden. Für dieses Projekt existiert jedoch weder eine Dokumentation noch eine offizielle Lizenz. Dafür sind die zwei erstellten Demos einzusehen, da es ein *Open-Source*-Projekt ist. Das Projekt kann in *JavaScript* genutzt werden. Der komplette Quellcode wurde noch nicht veröffentlicht, soll allerdings noch folgen. [93]

EmoPy

EmoPy ist ein *Open-Source*-Projekt zum Erkennen von Emotionen als ein Teil von ThoughtWorks¹. Die Berater unterstützen Karen Palmer² bei einem interaktiven Filmerlebnis, RIOT³. Die folgenden Emotionen können von der Software erkannt werden: Ekel, Wut, Überraschung, Trauer, Neutral, Freude, Angst. Das Projekt basiert auf einem künstlichen neuronalen Netz. Diese Modelle von neuronalen Netzen sind frei zugänglich, einfach zu recherchieren und in andere Projekte zu integrieren. Es nutzt zudem öffentlich zugängliche Datensätze. Das Projekt ist in *Python* geschrieben und kann entsprechend auch in *Python* genutzt werden. EmoPy hat eine Erklärung zum Aufsetzen der Entwicklungsumgebung, eine Installationsanleitung, eine Anleitung zum Starten der Tests und eine Anleitung zum Starten der Beispiele. Eine genauere Dokumentation ist allerdings nicht vorhanden. EmoPy hat eine AGPL-3.0 Lizenz. [100] [101] [103]

-
- 1 ThoughtWorks ist ein Unternehmen für Software-Beratung. [103] (<https://www.thoughtworks.com/>)
 - 2 Karen Palmer produziert Filme, die die Genres Film, Gaming, Kunst, Wissenschaft und Technologie vereinen. [104]
 - 3 RIOT ist ein Live-Action-Film von Karen Palmer, welcher KI und *Machine Learning* in Verbindung mit Gesichtserkennung nutzt. Die Zuschauer können durch den Film navigieren und die Geschichte des Films ändert sich je nach Reaktion der Zuschauer. [104] (<https://thoughtworksarts.io/projects/riot/>)

5.2.2. Ton

Die zweite Säule der Emotionserkennung ist der Eingabestrom Ton. Diese soll mithilfe einer Audiodatei die Emotion einer Person erkennen können. Im Folgenden werden ausgewählte Dienste und Projekte vorgestellt.

DELTA - A Deep learning Language Technology plAtform

DELTA ist eine *Open-Source*-Plattform, welche sich mit der *Verarbeitung natürlicher Sprache (NLP)* beschäftigt. Außerdem basiert sie auf Deep Learning. DELTA ist in *Python* implementiert mit der Unterstützung von TensorFlow. DELTA bietet neben der Emotionserkennung noch weitere Funktionen, wie Spracherkennung und Sprecherverifikation. Außerdem bietet sie im Bereich von *NLP* die Funktionen Textklassifikation, Eigennamenerkennung, Fragen und Antworten, Textzusammenfassung, etc. DELTA ist lizenziert durch die Apache-2.0 Lizenz. Auf der GitHub-Seite sind die Installationsanleitung und ein "Quick Start" vorhanden. Der "Quick Start" geht auf das Starten von Beispielen oder auch die Bereitstellung der Software ein. Außerdem hat DELTA eine Dokumentation, die unter anderem auch weitere Tutorials enthält. Die neuste Version ist am 16. Juli 2020 erschienen. [106] [107] [108] [110]

Speech Emotion Analyzer

Speech Emotion Analyzer ist ein *Open-Source*-Projekt, welches Emotionen anhand von Audiodateien erkennt. Folgende Emotionen können vom Programm erkannt werden: Neutral, Ruhe, Freude, Trauer, Wut, Angst, Ekel und Überraschung. Zusätzlich kann es bestimmen, ob die Person männlich oder weiblich ist. Das Projekt nutzt Deep Learning und *NLP* und ist in der Programmiersprache *Python* implementiert worden. Das Projekt ist mit der MIT-Lizenz lizenziert. Es liegen keinerlei Dokumentationen vor, lediglich sind ca. 40 beantwortete Fragen auf GitHub zu finden. [111]

5.2.3. Text

Die letzte Säule ist der Eingabestrom Text. Diese soll anhand von Text die Emotion einer Person bestimmen. Im Folgenden werden ausgewählte Dienste und Projekte vorgestellt.

IBM Watson Tone Analyzer

Der Tone Analyzer von IBM Watson¹ ist ein Dienst, der Emotionen bzw. den Ton eines Texts erkennen kann. Der Dienst kann sowohl kürzere Texte, wie zum Beispiel E-Mails, analysieren, als auch Chat-Konversationen, beispielsweise von einem Kundensupport. Der Text muss jedoch bei beiden Anwendungsgebieten in den Sprachen Englisch oder Französisch verfasst sein. Die folgenden Emotionen kann der Tone Analyzer erkennen: Wut, Angst, Freude und Trauer. Des Weiteren können folgende Töne erkannt werden: analytisch, selbstbewusst und zaghaft. Die Dokumentationen für den Service sind sehr ausführlich. Eine Demo² zum Testen steht bereits von IBM bereit. IBM bietet für den Tone Analyzer zwei verschiedene kostenfreie Versionen an. Der Lite-Plan stellt 2.500 kostenfreie API-Aufrufe im Monat zur Verfügung. Der Lite-Plan wird aber nach 30 Tagen ohne Nutzung gelöscht. Dieser Plan ist kostenlos. Der Standard-Plan bietet 1.000 kostenfreie API-Aufrufe im Monat an. Jeder weitere Aufruf kostet 0,007421€. [3] [113]

1 IBM Watson bietet Cloud-Dienste im Bereich KI an. [29] (<https://www.ibm.com/de-de/watson>)

2 IBM Watson Tone Analyzer (<https://tone-analyzer-demo.ng.bluemix.net/>)

5.2.4. Multimodal

Es existieren bereits multimodale Lösungen für die Emotionserkennung. Multimodale Lösungen nutzen mehrere Eingabeströme für die Emotionserkennung. Jedoch werden die Eingabeströme nicht miteinander genutzt und vereint, weshalb sie keine hybriden Lösungen darstellen.

Multimodal Emotion Recognition

Diese multimodale Emotionserkennung ist eine in *Python* implementierte *Open-Source-Web-Anwendung*, welche Emotionen mit den drei Eingabeströmen Bild, Ton und Text erkennen kann. Hierbei ist allerdings erwähnenswert, dass in der schon programmierten Anwendung keine hybride Lösung entstanden ist. Die Eingabeströme werden nicht miteinander in Verbindung gebracht, sondern sind einzeln nutzbar. Das Projekt wurde implementiert, um Emotionen von Bewerbern für einen Job erkennen zu können. [115]

Durch Bild und Ton können folgende Emotionen erkannt werden: Freude, Trauer, Angst, Wut, Überraschung, Neutralität und Ekel. Bei der Erkennung durch den Eingabestrom Text werden die folgenden Töne erkannt: Offenheit, Gewissenhaftigkeit, *Extraversion*, Verträglichkeit und *Neurotizismus*. Die Emotionserkennung nutzt für jeden Eingabestrom ein eigenes **CNN**, welches die Daten analysiert. [115]

Es existiert bereits eine GUI für diese Anwendung. In Abbildung 20 ist der Überblick der Anwendung zu erkennen. Auf dieser Seite kann ausgewählt werden, welches Interview durchgeführt werden soll. [115]

In Abbildung 21, 22 und 23 sind die Seiten der jeweiligen Interviews zu erkennen. Bei diesen können Dateien hochgeladen, aufgenommen oder Texte eingegeben werden. Diese Daten können mit einem Klick auf den Button "Start Analysis" analysiert werden. [115]

Das Projekt ist mit der Apache-2.0-Lizenz lizenziert. Eine Community exis-

tiert für dieses Projekt quasi nicht. Probleme und Fragen werden auf GitHub eher selten beantwortet. Jedoch werden alle paar Monate Fehler eliminiert. Eine ausführliche Dokumentation zur Implementierung dieses Projekts lässt sich als PDF abrufen. Auf GitHub lässt sich zusätzlich eine Installationsanleitung finden. [115]

Abb. 20: Überblick-Seite der multimodalen Emotionserkennung [115]

Abb. 21: Video-Interview-Seite der multimodalen Emotionserkennung [115]

Abb. 22: Audio-Interview-Seite der multimodalen Emotionserkennung [115]

Abb. 23: Text-Interview-Seite der multimodalen Emotionserkennung [115]

All diese verschiedenen Dienste und Projekte sind von ihrer Aufmachung sehr unterschiedlich. Doch viele Emotionserkennungen beruhen auf **Machine Learning** oder sogar Deep Learning. Ebenfalls fällt auf, dass die Emotionserkennung durch eine Texteingabe oft keine Emotionen als Ergebnis zurückgibt, sondern den generellen Ton eines Textes erfasst. Außerdem muss die Texteingabe einer bereits implementierten Lösung in einer bestimmten Sprache erfolgen, welche meist die Sprache Englisch ist. Ähnliches lässt sich auch bei der Emotionserkennung durch den Eingabestrom Ton erkennen. Hier fließt die Betonung in die Emotionserkennung mit ein. Diese ist jedoch in jeder Sprache unterschiedlich. Des Weiteren sind die zu erkennenden Emotionen immer auf eine bestimmte Anzahl begrenzt. Aus diesen Gründen werden im nächsten Kapitel verschiedene Algorithmen zur Emotionserkennung aufgezeigt. Somit kann eine hybride Lösung zur Emotionserkennung mit den Emotionen, die gebraucht werden, entstehen. Des Weiteren besteht die Möglichkeit, dass die Lösung später erweitert werden kann. [70]

5.3. Algorithmen

In diesem Abschnitt wird auf verschiedene Algorithmen zur Erkennung von Emotionen eingegangen. Hierbei wird die Funktionsweise der Algorithmen erläutert. Des Weiteren wird die Genauigkeit der Algorithmen genannt. Auch hier werden verschiedene Algorithmen zu den einzelnen Säulen aufgezeigt.

5.3.1. Bild

Die gesichtsbasierte Emotionserkennung ist die am häufigsten genutzte. Das liegt daran, dass die meisten Systeme mit Kameras oder Videomaterial arbeiten, unabhängig von ihrem Einsatzgebiet. Ein weiterer Faktor ist, dass die Gesichtsausdrücke unter anderem zu den wichtigsten Merkmalen beim Ausdruck von Emotionen zählen. Aus diesem Grund lassen sich hier viele

verschiedene Ansätze und Algorithmen finden. Dennoch ist dies kein einfach zu lösendes Problem, auch nicht mit Machine Learning. Menschen sind sehr unterschiedlich und unterscheiden sich auch darin, wie sie ihre Emotionen zeigen und ausdrücken. Zudem werden Algorithmen auch nur auf eine bestimmte Auswahl von Emotionen reduziert. Meist können diese Grundemotionen wie Wut, Trauer, Freude oder Angst erkennen. Manche Algorithmen sind aber auch auf Emotionen wie Überraschung, Ekel oder auf den neutralen Gesichtsausdruck ausgelegt. [69] [72]

Generell ist die **Bildvorverarbeitung** bei jedem Ansatz sehr ähnlich. Je nach gewünschter Genauigkeit des Algorithmus kann die Vorverarbeitung eines Bildes sehr zeitintensiv ausfallen. Hierbei müssen bzw. können mehrere Punkte beachtet werden. Das Umfeld spielt hierbei eine große Rolle. Wie sind die Lichtverhältnisse? Sollen Emotionen auch ohne zusätzliche Lichtquellen erkannt werden, zum Beispiel beim Einsatz im Smart Home? Sollen Emotionen von mehreren Personen auf einem Bild- oder Videomaterial zu erkennen sein? Sollen nur bestimmte Personen analysiert werden? Schauen die Personen immer direkt in die Kamera? Diese Liste an Fragen zur Umgebung kann sehr ausführlich fortgeführt werden, je nachdem welche Szenarien für den Einsatz der Emotionserkennung von Bedeutung sind. [69] [72]

Abb. 24: Facial Feature Point Detection und die daraus resultierende konvexe Hülle [122]

Der erste Schritt ist dann die Detektion des Gesichts. Hierbei werden Informationen, die nicht zum Gesicht gehören, eliminiert. Dies macht den nächsten Schritt effizienter und akkurater. Der nächste Schritt besteht darin, die Gesichtsmerkmale durch Punkte zu rekreieren. Zu den Gesichtsmerkmalen, zählen Augen, Augenbrauen, Nase, Mund und Kinn. Durch diese Punkte im Gesicht können Emotionen ausgedrückt werden. Dieser Schritt wird durch Facial Feature Point Detection (FFDP) realisiert. Hierbei wird ein bereits vorgefertigter Satz an Gesichtspunkten in ein Bild mit Gesicht eingebaut, wie auf der linken Seite in Abbildung 24 zu sehen ist. [72]

Durch zwei verschiedene Ansätze können die Gesichtsausdrücke erkannt werden. Ein Ansatz basiert auf der Verarbeitung von **statischen Bildern**. Bei diesem statisch-basierten Verfahren enthält ein Vektor nur Informationen über die Merkmale des Gesichts vom aktuellen Eingabebild. [72]

Der zweite Ansatz arbeitet mit **Bild-Sequenzen**. Dabei werden zusätzlich zeitliche Informationen von Bildern genutzt, um den Gesichtsausdruck zu erkennen. Dabei wird ein oder auch mehrere Zeitfenster genutzt. Bei automatisierten Systemen zur Erkennung von Gesichtsausdrücken können beide Ansätze genutzt werden. Somit können diese als Eingabe sowohl ein statisches Bild, als auch eine Bildsequenz empfangen und als Ausgaben geben diese den Gesichtsausdruck zurück. [72]

Außerdem lässt sich die Emotionserkennung mit Bildmaterial in konventionelle Ansätze und Ansätze mit Deep Learning unterteilen. Die konventionellen Ansätze basieren auf der Klassifikation von Merkmalen eines *Histogramms* orientierter *Gradienten* (HOG). Ein HOG bezeichnet die Verteilung von Steigungen innerhalb eines Bildes. Es wird dazu genutzt, um Objekte, in diesem Fall die Merkmale eines Gesichts, zu erkennen. Die Deep Learning Ansätze basieren auf Convolutional Neural Networks (CNN). Es werden im Folgenden zwei konventionelle Ansätze:

- Support Vector Machine (SVM)
- Multilayer Perceptron (MLP)

und drei Ansätze mit Machine Learning vorgestellt:

- Affdex CNN
- AlexNet CNN
- FER-CNN

Diese Ansätze wurden aus einer wissenschaftlichen Arbeit¹ ausgewählt, getestet und die Ergebnisse bereitgestellt. [72] [123]

Die konventionellen Ansätze werden in den folgenden Schritten durchgeführt. Die ersten Schritte bestehen darin, die Gesichtsmarkierungen im Gesicht (Augen, Augenbrauen, Nase, Mund, Kinn) zu erkennen und das Gesicht zu extrahieren. Diese beiden ersten Schritte werden an Graustufen-Bildern mithilfe von OpenFace² für MATLAB³ durchgeführt. Die Erkennung von den einzelnen Gesichtsmarkierungen wird durch einen Algorithmus in OpenFace durchgeführt. Dieser Algorithmus basiert auf CLNF (Constrained Local Neural Field) und CLM-Modellen (Constrained Local Model). **CLM-Modelle** sind bewährte Methoden zur Detektion von Gesichtsmarkierungen. **CLNF** ist ein Modell zur Detektion von Gesichtsmarkierungen, welches Verbesserungen in schlechten Lichtverhältnissen mit sich bringt und zudem lernfähig ist. Anhand der bereits extrahierten Gesichtsmarkierungen in Abbildung 24, können nicht-gesichtsbezogene Informationen aus dem Bild entfernt werden. Dazu wird eine binäre (schwarz/weiß) Maske (Abbildung 25, b) unter Verwendung einer konvexen Hülle erstellt, die die Gesichtspunkte umgibt. Die konvexe Hülle resultiert aus den vorher bestimmten Gesichtsmarkierungen, wie auch in Abbildung 24 zu erkennen, wobei die rechte Seite der Abbildung die konvexe Hülle darstellt.

-
- 1 Real-time Algorithms for Facial Emotion Recognition: A Comparison of Different Approaches (<https://ieeexplore.ieee.org/document/8587011>)
 - 2 OpenFace ist eine Erweiterung für MATLAB, die es ermöglicht Gesichter zu analysieren. [118] (<https://github.com/TadasBaltrusaitis/OpenFace>)
 - 3 MATLAB ist ein Programm, um mathematische Prozesse zu programmieren und grafisch darzustellen. [119] (<https://de.mathworks.com/products/matlab.html>)

Somit kann das Gesicht extrahiert werden, wie in Abbildung 25, c und d dargestellt. [72] [120] [121]

Abb. 25: Extraktion eines Gesichts mithilfe von FFPD [72]

Im nächsten Schritt werden die Merkmale im Gesicht extrahiert und klassifiziert. Dazu werden die Klassifizierungsmethoden SVM und MLP-KNN verwendet, um unterschiedliche Emotionen basierend auf HOG-Merkmalen aus verschiedenen Gesichtsbereichen zu unterscheiden. [72]

SVM ist eine Methode, um Daten zu analysieren und Objekten bestimmten Klassen zuzuordnen. Bei dieser Methode ist zunächst ein Training erforderlich. Dabei ist es wichtig, dass die Objekte bereits den Klassen zugeordnet sind. Hierbei werden die Gesichtsmarkierungen von OpenFace aus MATLAB den verschiedenen Emotionen zugeordnet. Die Methode übernimmt dann die Trennung der verschiedenen Klassen. Somit können neue Gesichtsmarkierungen den Klassen zugeordnet werden. [72] [124]

Im Folgenden werden die drei Machine-Learning-Ansätze vorgestellt. Jeder der Ansätze ist auf einem CNN aufgebaut. Der erste Ansatz besteht darin, ein bereits trainiertes **AlexNet CNN** zu nutzen. AlexNet wurde nach einem seiner Erfinder genannt. In einer wissenschaftlichen Arbeit¹ von 2012 wurde dieses CNN für einen Wettbewerb implementiert, welches 1,2 Millionen

1 ImageNet Classification with Deep Convolutional Neural Networks (<https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>)

Bild-Daten klassifiziert. Das CNN wies eine Fehlerrate von 15,3 % auf. Platz zwei hatte im Vergleich dazu eine Fehlerrate von 26,2 %. Damit hatte es die **geringste Fehlerrate** in dem gesamten Wettbewerb. AlexNet sorgte dafür, dass in den folgenden Jahren CNNs entwickelt wurden, deren Fehlerrate an die eines Menschen herankommen. Es besteht aus 650.000 Neuronen, fünf Faltungsschichten und drei vollständig vermaschten Schichten. [72] [126] [127]

Für den Ansatz wurde das AlexNet CNN für den genauen Anwendungsfall verbessert. Dazu wurde die Methode **Transfer Learning** verwendet. Transfer Learning bezeichnet hier die Nutzung eines bereits trainierten CNNs, welches für eine andere Aufgabe wiederverwendet werden soll. Diese optimiert die bereits erlernten Parameter des CNN und löst somit das Klassifizierungsproblem für Emotionen. [72] [128]

Um die vier Grundemotionen darzustellen wurden die Ausgänge der letzten vollständig vermaschten Schicht auf vier reduziert. Somit lassen sich die vier Emotionen Freude, Trauer, Wut und Angst abbilden. Die beiden vorletzten Schichten wurden ebenfalls ersetzt. Dabei wurden die ursprünglichen Gewichte der ersetzten Schichten zufällig ausgewählt. In dem gesamten Prozess wurden nicht die einzelnen ersetzten Schichten trainiert, sondern das gesamte Netzwerk. Aus diesem Grund ist es wichtig, dass die globale Lernrate auf einen kleinen Wert einzustellen. Somit wurde sichergestellt, dass das Lernen von den ersetzten, vollständig vermaschten Schichten in einem schnellen Prozess erfolgt. Wohingegen die Lernrate der einzelnen Schichten auf einen höheren Wert gesetzt wird, sodass das Lernen der übertragenden Schichten in einem langsamen Prozess erfolgt. [72]

Bei einem Trainingsprozess kann es dazu kommen, dass das CNN neue Daten, die nicht im Datensatz des Trainings enthalten sind, nicht klassifizieren kann. Um diesem vorzubeugen, werden neue Daten erzeugt. Aus jedem Bild aus dem Trainingsdatensatz werden zehn weitere erzeugt. [72] [129]

Für die Feinabstimmung vom AlexNet wurden 70 % der Bilder-Datensätze genutzt. Die restlichen 30 % wurden zum Testen genutzt. [72]

Der zweite Deep-Learning-Ansatz wird **FER-CNN** genannt. Dieser wurde von Studenten der Technischen Universität in Delft entwickelt. Der Name des Ansatzes bezieht sich auf den für das Training genutzten Datensatz. FER ist die Abkürzung für "Facial Expression Recognition". Der genutzte Datensatz ist FERC-2013. Das CNN besteht aus drei Faltungsschichten und einer vollständig vermaschten Schicht. Die Eingabe des CNNs ist das extrahierte Gesicht. Das Gesicht wird mithilfe des **Viola-Jones Algorithmus** extrahiert. Der Algorithmus ist eine Echtzeitverarbeitung für die Detektion eines Gesichts. Er ist schnell und genau in der Verarbeitung. Der Algorithmus basiert auf dem **Haar-Cascade-Algorithmus** und verwendet ein *integrales Bild*, um die **Echtzeitverarbeitung** zu ermöglichen. [69] [72]

Die erste Faltungsschicht gibt eine 44x44x64 Merkmalsmatrix zurück. Nach der ersten Pooling-Schicht wird die Merkmalsmatrix auf 22x22 reduziert. Die zweite Faltungsschicht gibt eine 18x18x64 Merkmalsmatrix zurück. [72]

Der dritte und letzte Ansatz wird mit dem **Affdex¹ CNN** durchgeführt. Die Affdex SDK besteht aus einer API, die sieben Emotionen, 20 Gesichtsausdrücke und 13 Emojis erkennen kann. In diesem Beispiel wurde die SDK als Unity²-Paket genutzt. Die mit Affdex erkannten Emotionen werden an komplexen Datensätzen trainiert und getestet. Der Trainingsdatensatz bestand aus Hunderttausenden von Gesichtsbildern, welche aus 3,2 Millionen Videos stammen. Die Daten stammen außerdem aus mehr als 75 verschiedenen Ländern und repräsentieren reale und spontane Gesichtsausdrücke. Diese wurden unter schwierigen Bedingungen aufgenommen. Damit weisen sie unterschiedliche Beleuchtungen, Kopfbewegungen, Unterschiede in ethnischer Zugehörigkeit, Alter, Geschlecht, Gesichtshaarung und Körperschmuck auf. Aus diesen Gründen ist dieser Ansatz der komplexeste der bisher vorgestellten. [72] [134]

1 Affdex: Eine plattformübergreifende SDK zur Erkennung mehrerer Gesichtsausdrücke in Echtzeit. (<https://dl.acm.org/doi/10.1145/2851581.2890247>)

2 Unity ist eine Echtzeit-3D-Entwicklungsplattform für Spiele. (<https://unity.com/>)

In einem kleinen Test mit acht Personen, stellte sich das **Affdex CNN** mit einer Genauigkeit von ungefähr 85 % als die **genaueste Anwendung** heraus. Die zweithöchste Genauigkeit von 76 % hatte das verbesserte AlexNet CNN. Mit ungefähr 63 % Genauigkeit folgt der Algorithmus mit der SVM Klassifikation von HOG Merkmalen. Die niedrigste Genauigkeit zeigte die MLP Klassifikation von HOG Merkmalen mit 56 % und der FER-CNN mit nur 55%. Mit dem Test sollte gezeigt werden, inwiefern es Sinn macht, einen eigenen Algorithmus zu entwickeln bzw. ein eigenes Netzwerk aufzusetzen, anstatt ein kommerzielles Programm zu nutzen. [72]

Die verschiedenen Ansätze zeigen teilweise große Unterschiede und sind daher schwer vergleichbar. Die genauesten Ansätze sind die, die auf Machine Learning beruhen. Jedoch ist die Aussagekraft des Tests fragwürdig, da dieser mit nur acht Personen durchgeführt worden ist.

5.3.2. Ton

Die Sprache ist für den Menschen der natürlichste Weg zu kommunizieren. Die übliche tonbasierte Emotionserkennung wird in drei Schritten durchgeführt. Mit dem Sprachsignal wird eine Vorverarbeitung durchgeführt, anschließend Merkmale extrahiert und zuletzt ein Klassifikator angewendet, um die Emotion zu bestimmen. Ansätze, die auf Machine Learning beruhen, geben die extrahierten Merkmale in ein neuronales Netz weiter, um die Emotion zu bestimmen. Die tonbasierte Emotionserkennung mit Machine Learning hat den Vorteil, dass die Erkennung effizienter ist. [135]

Ein Ansatz ist der **Ende-zu-Ende Emotionserkennung Algorithmus**. Ende-zu-Ende bezeichnet beim Machine Learning, dass die Ausgabe von der Eingabe abhängig ist. Der Ansatz nutzt den **Backpropagation-Algorithmus**, welcher die weitverbreitetste Lernmethode für neuronale Netze ist.

Dieser Algorithmus bestimmt das Minimum der Fehlerfunktion eines bestimmten Problems. Die Lernrate wird ständig durch den Algorithmus angepasst, um die Gewichtungen innerhalb des neuronalen Netzes zu ändern. Der gesamte Ansatz wurde bereits mehrmals umgesetzt. Dabei ist der Grundaufbau jedes Ansatzes gleich. Jedoch lassen sich **Unterschiede in der Eingabe des Sprachsignals** und in der Wahl des neuronalen Netzes erkennen. Die verschiedenen Arten der Eingaben werden Prozesse genannt, welche in Abbildung 26 dargestellt werden. Alle drei Prozesse arbeiten mit dem Sprachsignal und einem neuronalen Netz, um eine Emotion zu erhalten. In dem ersten Prozess wird das Sprachsignal unverändert in das neuronale Netz gegeben. In dem zweiten Prozess wird das Signal in ein *Spektrogramm* umgewandelt und anschließend in das neuronale Netz gegeben. Bei dem dritten Prozess wird das Signal in ihrer Frequenz oder in ihrem Dynamikbereich verändert und daraufhin in das neuronale Netz gegeben. [112, S. 87] [114, S. 149] [135]

1. Prozess:

2. Prozess:

3. Prozess:

Abb. 26: Prozesse der Ende-zu-Ende Emotionserkennung. Abbildung nach [135]

Das neuronale Netz ist bei den Ansätzen meist nur ein **CNN** oder ein CNN in Verbindung mit einem anderen neuronalen Netz, wie ein RNN, LSTM oder DNN. [135]

Der Ende-zu-Ende Emotionserkennung Algorithmus hat den Vorteil, dass die **Vorverarbeitung reduziert** wird. Dazu zählen die Vorverarbeitung von Sprachsignalen und die Extraktion von Merkmalen. Des Weiteren können diese komplexere Aufgaben mit einer höheren Genauigkeit lösen [135]

Generell birgt die tonbasierte Emotionserkennung jedoch viele Schwierigkeiten. Die größte Hürde war schon immer die Extraktion von Informationen bei Hintergrundgeräuschen. Sehr gut hingegen funktioniert sie schon in ruhigen Umgebungen. Hintergrundgeräusche sind jedoch sehr zufällig, können sich ständig ändern und unterscheiden sich auch ständig in zeitlicher Hinsicht. Aber auch die unterschiedliche Sprechweise von Menschen und auch die unterschiedliche Betonung in verschiedenen Sprachen und Kulturen können Hindernisse darstellen. Die meisten Emotionserkennungen auf Basis von Ton-Daten fand in den Sprachen Englisch, Japanisch, Deutsch und Spanisch Anwendung. Generell hat sich die tonbasierte Emotionserkennung viel damit beschäftigt, welche emotionalen Informationen und Charakteristiken in einem Sprachsignal enthalten sein können. [70]

5.3.3. Text

Die textbasierte Emotionserkennung ist die schwierigste der drei vorgestellten Eingabeströme. Studien der textbasierten Emotionserkennung nutzen derzeit vor allem *NLP* und Machine Learning. Es werden solch komplexe Methoden verwendet, da es in Texten meist einen gesamten Kontext zu verstehen gibt. Dieser wird für spezifische Informationen gebraucht, um den gesamten Text verstehen zu können. Gebraucht wird eine textbasierte Emotionserkennung zum Beispiel in einer Analyse von Rezensionen oder

für interaktive *Chatbots*. Im Allgemeinen lässt sich sagen, dass ein Text, eine Geschichte oder eine Unterhaltung immer drei Elemente umfasst. Diese drei Elemente sind der Charakter, das Ereignis und der Rahmen, wie zum Beispiel der Ort und die Zeit. Diese drei Elemente sind wichtige Informationen in einem Text für die Emotionserkennung. Aber auch die Textgenerierung und das Textverständnis spielen eine Rolle. [27]

Die textbasierte Emotionserkennung hat sich von einer sehr groben zu einer feinen Erkennung entwickelt. Eine grobe Erkennung bezeichnet die Klassifizierung von Emotionen in nur zwei verschiedene: positiv und negativ. Die feine Erkennung von Emotionen kann Emotionen in viele verschieden klassifizieren. Meist werden Emotionen in vier verschiedene unterteilt: Freude, Trauer, Wut und Angst. [27]

In diesem Beispiel zur textbasierten Emotionserkennung werden acht Grundemotionen, basierend auf dem Rad der Emotionen von Plutchik, genutzt: Wut, Erwartung, Ekel, Angst, Freude, Vertrauen, Trauer und Überraschung. Hierbei werden jedoch nur die primären Emotionen genutzt. Auf die Intensitätsstufen oder die Kombinationen der Emotionen wird nicht eingegangen. [27]

Für die textbasierte Emotionserkennung gibt es zwei grundlegende Methoden: die schlüsselwortbasierte Methode und die lernbasierte Methode. Dabei ist die **schlüsselwortbasierte Methode** die am meisten und einfachsten zu nutzende Methode in der textbasierten Emotionserkennung. Zuerst werden Wörter, die Emotionen ausdrücken können, in einem Text herausgesucht. Mit diesem Hinweis, selbstdefinierten Regeln und einem bestimmten Vokabular werden die Emotionen in einem Satz bestimmt. Diese Methode benutzt ein Wörterbuch, in dem eine große Anzahl an Wörtern mit Emotionen verknüpft werden. [27]

Die **lernbasierte Methode** basiert auf einem trainierten Modell mit einer großen Anzahl an Daten. Dieses soll mithilfe von Testdaten Emotionen klassifizieren können. Machine-Learning-Algorithmen, wie zum Beispiel

SVM oder *Entscheidungsbäume*, werden für die Klassifikation von Emotionen genutzt. Für die Erkennung von Emotionen innerhalb eines Texts werden Deep-Learning-Algorithmen, zum Beispiel ein CNN oder LSTM, genutzt. [27]

Eine neue Technik für die Klassifizierung von Emotionen in Texten ist das **Emotion Embedding Model**, welches Word Embedding nutzt. Word Embedding ist ein Algorithmus, der die Wörter durch Vektoren ersetzt, da unveränderte Wörter nicht in ein neuronales Netz oder in einen anderen Machine-Learning-Algorithmus gegeben werden können. Wörter, die Ähnliches aussagen, werden hierbei in Zusammenhang gebracht. Für die Emotionserkennung bedeutet es, dass Wörtern Emotionen zugeordnet werden. Wörter, die ähnliche Emotionen ausdrücken, werden näher zueinander gelegt während Wörter, die andere Emotionen ausdrücken, weiter entfernt gelegt werden. [27] [46]

Im Folgenden wird ein Ansatz erläutert, welcher das Emotion Embedding Modell nutzt, um in Texten verschiedene Emotionen erkennen zu können. Dieser Ansatz wurde in einer wissenschaftlichen Arbeit¹ implementiert. Der Ansatz ist in Abbildung 27 dargestellt. [27]

Abb. 27: Textbasierte Emotionserkennung mit einem Emotion Embedding Model [27]

1 Emotion Recognition from Text Stories Using an Emotion Embedding Model (<https://ieeexplore.ieee.org/document/9070506>)

Im ersten Schritt mussten dafür Daten gesammelt werden. Hierfür wurden Daten von Twitter¹ genutzt. Dabei wurden Tweets² gesucht, hinter denen jeweils ein *Hashtag* mit einem Wort, welches eine Emotion ausdrückt, geschrieben wurde. Solch ein Tweet kann wie folgt aussehen: "Ich habe mit meinem Freund Schluss gemacht. #traurig". Hierbei wurden, wie oben erwähnt, die acht grundlegenden Emotionen, basierend auf Plutchiks Rad der Emotionen, genutzt. [27]

Im zweiten Schritt wurde das Emotion Embedding Modell konstruiert. Hierfür wurde ein **CNN** genutzt, um die Emotionen zu klassifizieren. Bei diesem Lernprozess wird eine Embedding-Schicht geschaffen, um den Text zu vektorisieren. Diese Schicht wird genutzt, um Emotionen zu detektieren und zu klassifizieren. [27]

Im nächsten Schritt werden **emotionale Wörter** aus den Texten **extrahiert**. Hierfür wird ein Datensatz verwendet, welcher verschiedene Geschichten enthält, welche jeweils fünf Sätze enthalten. Mit dem NLTK VADER Sentiment Analyzer³ wurden die emotionalen Wörter innerhalb der Texte detektiert und bewertet. Dieser bewertet Wörter mit vier verschiedenen Werten: positiv, negativ, neutral und die Verbindung dieser drei Werte. Die Verbindung kann Werte im Bereich von -1 bis +1 annehmen. Dabei bedeutet -1 die höchste Negativität und +1 die höchste Positivität. Der höchste absolute Wert der Verbindung wird genutzt, um das emotionale Wort innerhalb eines Satzes zu erkennen. Dieser drückt entweder eine sehr positive oder eine sehr negative Bedeutung für ein Wort aus. Das bedeutet wiederum, dass das emotionalste Wort in einem Satz gefunden wurde. In Abbildung 28 mit einem Beispiel wird gezeigt, wie eine solche Bewertung eines einzelnen Satzes aussehen kann. [27]

1 Twitter ist ein soziales Netzwerk. (<https://twitter.com/>)

2 Ein Tweet ist eine Statusmeldung der Plattform Twitter. [44]

3 Natural Language Toolkit (NLTK) ist eine Plattform für *Python*, welche es ermöglicht mit von Menschen geschriebenen Daten zu arbeiten. (<http://www.nltk.org/>) Der Sentiment Analyzer ist ein Teil des NLTK und analysiert die Töne innerhalb eines Textes. (<http://www.nltk.org/howto/sentiment.html>) [42]

"one day a guest made him very **angry**"
[0,0; 0,0; 0,0; 0,0; 0,0; 0,0; 0,0; 0,121; **-0,5106**]

Abb. 28: Beispiel zum Finden des emotionalen Wortes eines Satzes. Abbildung nach [27]

In diesem Beispiel lässt sich gut erkennen, dass nur zwei Wörter überhaupt emotionale Tendenzen aufweisen: "very" und "angry". Da "angry" allerdings den höheren absoluten Wert besitzt, wird es als das emotionale Wort des Satzes ausgewählt. [27]

Im letzten Schritt werden die Emotionen von den Texten mithilfe des Emotion Embedding Model bestimmt. Somit wird das vorher bestimmte emotionale Wort eines Satzes einer Emotion zugeordnet. Dazu werden ähnliche Wörter im Emotion Embedding Modell genutzt. Bei dem Beispiel in Abbildung 28 ist "angry" das emotionale Wort und würde Wut als Emotion zugeschrieben werden. [27]

Der gesamte Prozess wurde an über 130.000 Sätzen durchgeführt. Für den Vergleich, welche Emotionen der Satz tatsächlich aufzeigt, wurden 120 Sätze (15 für jede Emotion) von Studenten mit einer Emotion gekennzeichnet. Die Ergebnisse zeigten eine Genauigkeit zwischen 36,7 % und 73,3 % bei der Erkennung der einzelnen Emotionen, wie in Tabelle 2 erkennbar ist. Der größte Schwachpunkt an dieser Methode liegt in der Wahl von nur einem emotionalen Wort pro Satz. Jedoch können Wörter auch verneint werden. Ein Satz könnte

Tabelle 2: Genauigkeit der zu erkennenden Emotionen. Tabelle nach [27]

Emotion	Genauigkeit
Wut	0,367
Antizipation	0,567
Ekel	0,55
Angst	0,483
Freude	0,733
Vertrauen	0,517
Trauer	0,45
Überraschung	0,433

zum Beispiel lauten: "Er war nicht glücklich darüber.". Hierbei würde das Wort "glücklich" als emotionales Wort festgelegt und die Emotion Freude erkannt werden. Allerdings wird das Wort "glücklich" verneint und sollte

nicht der Emotion Freude zugeordnet werden. Auch Sätze, die in einem Nebensatz etwas Negatives erwähnen, wie zum Beispiel: "Er war entschlossen zu spielen, wusste aber, dass er nachts lernen musste.". Hierbei würde das Wort "spielen" als emotionales Wort gewählt und die Emotion als Freude erkannt werden. Dabei sagt der Satz durch den Nebensatz keine Freude mehr aus. [27]

Mit den vielen Ansätzen lässt sich erahnen wie viele Möglichkeiten es in der Emotionserkennung gibt. Verschiedene Algorithmen und verschiedene Machine-Learning-Ansätze liefern viele verschiedene Ergebnisse. Viele dieser Ansätze liefern schon eine sehr zuverlässige Emotionserkennung. Jedoch scheint es die perfekte Erkennung noch nicht zu geben. Die genauesten Ergebnisse bringt die gesichts-basierte Emotionserkennung hervor. Dies kann vor allem daran liegen, dass es die meisten Forschungen und somit auch die meisten Informationen dazu gibt, Emotionen aus einem Gesichtsausdruck zu erkennen. Es existieren keine wissenschaftlichen Arbeiten, welche den direkt Einfluss von Alter oder Geschlecht auf die Emotionserkennung aufzeigen. Ein genaueres Ergebnis kann jedoch durch das Zusammenspiel von mehreren Informationen entstehen. Aus diesem Grund wird im folgenden Kapitel ein hybrider Algorithmus konzipiert.

6. Konzept

In diesem Kapitel wird ein Konzept für eine prototypische, hybride Emotionserkennung entwickelt. Hierbei werden die bereits vorgestellten Dienste, Programme und Algorithmen genutzt.

6.1. Abgrenzung

Ein hybrider Algorithmus zur Emotionserkennung ist ein komplexes Thema. Um dem Stand der Technik gerecht zu werden, müsste ein neuronales Netz aufgesetzt werden. Dieses ist sehr aufwändig zu implementieren, liefert jedoch sehr genaue Ergebnisse. Entweder müsste das neuronale Netz auf mehrere Eingaben ausgelegt sein, um mindestens zwei verschiedene Eingabeströme zu akzeptieren, oder mindestens zwei neuronale Netze implementiert werden, wobei eins für jeden Eingabestrom. Mit einem dieser Ansätze würde eine hybride Lösung entstehen. Beim Anlernen der neuronalen Netze müssen die Abgrenzungen festgelegt werden. Dabei ist es wichtig zu wissen, unter welchen Umständen die Emotionserkennung funktionieren soll. Soll die Emotionserkennung in verschiedenen Lichtverhältnissen, sowie in Dunkelheit, funktionieren? Soll sie mehrere Personen erkennen können? Soll sie mit vielen lauten und leisen Hintergrundgeräuschen funktionieren? All diese Faktoren müssten in das Konzept mit einfließen.

Dennoch würde mit einer eigenen Implementierung das Konzept nahezu perfekt auf den Roboter ROS-E angepasst werden. Ebenso könnten **neuronale Netze** mit eigenen Daten antrainiert werden, um somit höhere Genauigkeiten für die entsprechenden Einsatzszenarien zu liefern. Auch die Programmiersprache und die GUI könnten passend ausgewählt und gestaltet werden, je nachdem, welche als effizienter erachtet oder besser in das Gesamtkonzept passen würde. Dennoch übersteigt dies den Rahmen einer prototypischen Implementierung und wird deshalb **nicht weiter in Betracht gezogen**.

Ebenfalls wird von Techniken abgesehen, die sich außerhalb des Möglichen dieser Thesis befinden. Deshalb wird sich bei der Auswahl der Dienste und *Open-Source*-Projekte auf die drei Grundpfeiler der Emotionen konzentriert. Dennoch können durch genutzte Dienste zusätzliche Daten mit einfließen, wie zum Beispiel das geschätzte Alter einer Person. Microsoft Azure bietet unter anderem auch die Schätzung des Alters einer Person an, welches dann zusätzlich zur Emotionserkennung genutzt werden kann.

Des Weiteren wird vorerst **von Live-Daten abgesehen**. Für die prototypische Anwendung werden vorgefertigte oder bereits existierende Daten oder Datensätze genutzt. Dies gilt für Tondateien, Bilder, Videos und Texte.

Bei einer hybriden Anwendung ist es wichtig, wie die einzelnen Komponenten miteinander zusammenspielen können und müssen. Dazu kann man sich die zwei folgenden fiktiven **Anwendungsszenarien** zur Hilfe ziehen.

Martin Müller ist 82 Jahre alt und möchte einen neuen Fernseher im Elektronikfachhandel kaufen. Martin scheint sich eine schlechte Zeit für die Mitarbeiter ausgesucht zu haben, denn diese sind in der Mittagspause oder reden bereits mit einem Kunden. Er entschließt sich dazu zum ersten mal den Roboter Sam nach seinem Anliegen zu fragen. Roboter Sam bietet ihm die Möglichkeit an mit ihm zu reden oder sein Anliegen auf seinem Touch-Display einzugeben. Da Martin aber schon Schwierigkeiten beim Tippen auf seinem Smartphone hat, entscheidet er sich für die altmodische Variante. Er erklärt Sam, dass er einen neuen Fernseher sucht. Sam erklärt ihm, wo er diesen im Markt findet und stellt ihm weitere Fragen dazu, was er sich beim Kauf eines neuen Fernsehers wünschen würde. Schnell kann sich Martin auf diese Art und Weise für einen neuen Fernseher entscheiden und kauft diesen anschließend.

Dirk Kaufmann ist 36 Jahre alt und möchte einen neuen Fernseher im Elektronikfachhandel kaufen. Dirk war schon öfter in diesem Markt, weshalb er schon öfter mit dem Roboter Sam interagiert hat. Da er damit nur gute Erfahrungen gesammelt hat und generell nicht unbedingt gern mit Menschen redet, entschließt er sich dazu den Roboter Sam nach seinem Anliegen zu fragen. Roboter Sam bietet ihm die Möglichkeit mit ihm zu reden oder sein Anliegen auf seinem Touch-Display

einzugeben. Dirk tippt sein Anliegen in das Textfeld auf dem Display ein. Sam antwortet ihm sprachlich und in Textform über das Tablet, und liefert ihm verschiedene Informationen zu Fernsehern. Dirk lässt sich zu den Fernsehern begleiten und stellt Sam danach noch einige Fragen über das Display, um verschiedene Fernseher miteinander vergleichen zu können. Auf diese Art und Weise hat sich Dirk für einen neuen Fernseher entschieden und kauft diesen anschließend.

Diese beiden Szenarien ähneln sich, doch der bedeutende Unterschied liegt in der Interaktionsart zwischen dem Menschen und dem Roboter. Die Interaktion zwischen Mensch und Roboter kann beinahe alle möglichen Eingaben umfassen. Für die Emotionserkennung ist es jedoch **unlogisch** eine **Emotionserkennung mit allen Eingabeströmen durchzuführen**. Das zeigen die beiden Szenarien. Falls eine Eingabe durch den Eingabestrom Ton erfolgt ist, wird keine Eingabe mehr durch den Eingabestrom Text erfolgen. Genauso wenig wird eine Eingabe durch den Eingabestrom Ton erfolgen, falls bereits eine Eingabe durch den Eingabestrom Text erfolgt ist. Der Eingabestrom Bild kann jedoch jederzeit zusätzlich zur Emotionserkennung genutzt werden. Hierbei ist es wichtig zu wissen, dass während der Eingabe eines Textes, die Person nicht direkt in die Kamera schauen wird. Je nach Blickwinkel der Person und Lage der Kamera kann es möglich sein, dass das Gesicht nicht korrekt erfasst wird. Im schlimmsten Fall kann es sein, dass nur der Oberkopf der Person zu sehen sein.

Hybride Emotionserkennung muss deshalb immer auf ihre Situation angepasst werden. Sie kann nur **eine aktive Eingabe** empfangen, aber **mehrere passive Eingaben**. Aktive Eingaben bezeichnen jene, die aktiv vom Nutzer eingegeben werden müssen, wie zum Beispiel eine Eingabe durch Ton oder Text. Eine passive Eingabe kann stets während der Anwendung erfasst werden, ohne eine direkte Eingabe durch den Nutzer, beispielsweise durch ein Bild.

Aus diesem Grund werden nur jeweils zwei Eingabeströme gleichzeitig betrachtet. Falls eine Texteingabe stattfindet, soll die Anwendung die Emotionserkennung über die Eingaben Text und Bild durchführen. Wenn in der Anwendung eine Toneingabe stattfindet, soll die Anwendung eine Emotionserkennung über die Eingaben Audio und Bild durchführen.

Zwei verschiedene Konzepte wären aufgrund dieser Abgrenzungen denkbar.

Zuallererst ist es möglich das *Open-Source*-Projekt "Multimodal Emotion Recognition" in ein hybrides Projekt umzuwandeln. Da hier schon drei verschiedene Faktoren zur Emotionserkennung genutzt werden, ist das ein guter und einfacher Start, um schnell Ergebnisse zu erzielen. Es existiert auch bereits eine GUI, die das spätere Bedienen vereinfacht und das Verständnis der durchgeführten Aktionen verbessert. Hierbei muss eine multimodale in eine hybride Anwendung umgewandelt werden.

Für ein zweites Konzept können die bereits vorgestellten Dienste und Programme aus Kapitel 5.3 zu einer hybriden Anwendung zusammengefügt werden. Hier besteht ebenfalls der Vorteil, dass die meisten Komponenten der Anwendung auf die entsprechenden Einsatzszenarien angepasst werden können.

6.2. Konzept 1

Das erste Konzept besteht darin, das *Open-Source*-Projekt "**Multimodal Emotion Recognition**" zu nutzen und dieses zu einer hybriden Anwendung zu ändern. Im Folgenden werden die Schritte beschrieben, um dies umzusetzen.

Die **GUI** kann dahingehend verändert werden, dass es statt drei verschiedenen Auswahlmöglichkeiten nur zwei gibt. So könnte eine Auswahlmöglichkeit zur Eingabe von Audioaufnahmen und eine Auswahlmöglichkeit zur Eingabe von Texten erzeugt werden. Das *Mock-Up* der neuen GUI ist in Abbildung 29 dargestellt. Die GUI der jeweiligen Eingabeflächen kann beibehalten werden. An der Funktionsweise der Anwendung müssen jedoch mehrere Anpassungen vorgenommen werden.

Abb. 29: *Mock-Up* der GUI für Konzept 1

Auswertung Bild und Audio

Bei der Auswertung einer Audiodatei muss zusätzlich eine Videodatei ausgewertet werden. Hierbei besteht die Möglichkeit während der Aufnahme der Audiodatei ebenfalls eine Videoaufnahme zu starten, um diese auszuwerten. Anderenfalls kann auch eine bereits erstellte Videodatei genutzt werden. Im Anschluss müssen beide Dateien separat voneinander ausgewertet werden. Beide werden jeweils ein Ergebnis aus den sechs verschiedenen Emotionen: Wut, Freude, Angst, Trauer, Überraschung und Ekel liefern. Sobald beide Ergebnisse dieselbe Emotion ergeben, wird diese Emotion auch bestimmt. Bei unterschiedlichen Ergebnissen müssen die jeweiligen Prozentzahlen der einzelnen Emotionen in einer anderen Art und Weise verwendet werden. Hier kann die Strategie herangezogen werden, bei der die Prozente als Punktzahlen gesehen werden. Somit können die einzelnen Punkte der Emotionen zusammengezählt werden, um die Emotion mit der höchsten Wahrscheinlichkeit zu ermitteln. Ein Beispiel dazu ist in Tabelle 3 dargestellt. Hier ist zu erkennen, dass die Prozente exakt als Punktzahlen übernommen werden. Die fett markierten Zahlen kennzeichnen die Ergebnisse der jeweiligen Emotionserkennungen. Die tonbasierte Emotionserkennung erkennt die Emotion Trauer. Die bildbasierte Emotionserkennung gibt jedoch die Emotion Ekel zurück.

Nachdem alle Werte addiert wurden, wird das höchste Ergebnis die Emotion Trauer zurückgeben.

Tabelle 3: Beispiel einer Auswertung zur Bestimmung von Emotionen bei unterschiedlichen willkürlichen Ergebnissen.

Emotion	Ton in %	Punktzahl Ton	Bild in %	Punktzahl Bild	Ergebnis
Wut	17	17	5	5	44
Freude	3	3	1	1	8
Angst	2	2	4	4	12
Trauer	40	40	36	36	152
Überraschung	11	11	12	12	46
Ekel	27	27	42	42	138

Auswertung Bild und Text

Bei der Auswertung eines Textes muss ein ähnliches Prinzip verfolgt werden. Hierbei wird ein Text eingegeben oder eine Video-Datei hochgeladen. Auch hier kann eine Videoaufnahme gestartet oder eine bereits hinterlegte Videodatei zu Testzwecken genutzt werden. Da die Auswertung eines Textes keine Emotionen zurückgeben kann, sondern die Big Five Persönlichkeitsfaktoren zurückgibt, muss eine andere Auswertung erfolgen. Die Big Five Persönlichkeitsfaktoren sind: *Neurotizismus*, *Extraversion*, *Offenheit*, *Verträglichkeit* und *Gewissenhaftigkeit*. Diese können nicht miteinander vereint werden. Aus diesem Grund werden bei dieser Erkennung eine Emotion und ein Ton ausgegeben. [23]

Sobald eine der beiden Erkennungen keine Emotion ermittelt, wird die jeweils andere Erkennung bei erkannter Emotion das Ergebnis zurückgeben. Falls keine der Beiden eine Emotion erkennt, sollte angezeigt werden, dass keine Emotion erkannt worden ist.

Die Anwendung basiert auf *Python*, *TensorFlow* und *OpenCV* und kann somit auf ROS-E genutzt werden. Es muss sichergestellt werden, dass mindestens *Python* 3.4 installiert ist. Des Weiteren müssen alle benötigten *Python*-Module installiert werden. Diese befinden sich in der Datei "requirements.txt". Module, wie Tensor-

Flow und OpenCV, sind bereits auf ROS-E installiert.

Für die Nutzung von Live-Daten muss innerhalb der Anwendung der Zugriff auf die Kamera und das Mikrofon von ROS-E implementiert werden.

Mit diesem Konzept ist es möglich CNNs für die hybride Emotionserkennung zu nutzen. Das Programm bietet Genauigkeiten von bis zu 85 % für die einzelnen Komponenten der Emotionserkennung. Somit stellt dieses Konzept eine einfache Möglichkeit zum Testen der hybriden Emotionserkennung dar.

6.3. Konzept 2

Das zweite Konzept setzt sich aus **mehreren Diensten und Programmen** zusammen, um eine hybride Anwendung zu implementieren. Hierfür werden der Dienst von **Microsoft Azure**, die Plattform **DELTA** und der Dienst von **IBM Watson** genutzt. Diese wurden aus den folgenden Gründen gewählt.

DELTA besitzt Dokumentationen und Tutorials, ist in *Python* implementiert und nutzt TensorFlow. Die Dienste von Microsoft Azure und IBM Watson lassen sich problemlos in die DELTA-Software integrieren. Zusätzlich sind diese Dienste von bekannten Firmen, weshalb davon auszugehen ist, dass sie nahezu perfekte Ergebnisse liefern.

Zuerst muss die Installation von DELTA durchgeführt werden. Dafür sollte sichergestellt werden, dass *Python* 3.6 installiert ist. Zur weiteren Installation muss die Anleitung auf GitHub¹ oder die Anleitung in der Dokumentation² genutzt werden. Für die Version, in der Sprachsignale verarbeitet werden können, muss zunächst Conda³ installiert werden. Anschließend muss Kaldi⁴ installiert werden. Zuletzt muss DELTA installiert werden und das Modell trainiert werden.

Nach der Installation von DELTA sollten jeweils ein Konto bei IBM Watson und Microsoft Azure angelegt werden. Somit sind die Schlüssel und URL zugänglich und die Dienste können genutzt werden.

Daraufhin muss eine GUI implementiert werden. Diese kann einfach gehalten werden, indem sie mit Tkinter aufgesetzt wird. Tkinter ist ein bereits in *Python* integrier-

1 GitHub DELTA Installationsanleitung: <https://github.com/didi/delta>

2 DELTA Dokumentation: https://delta-didi.readthedocs.io/en/latest/installation/install_from_source.html

3 Conda ist eine Verwaltung für Pakete und deren Abhängigkeiten für viele Programmiersprachen. <https://docs.conda.io/en/latest/> [136]

4 Kaldi ist ein Toolkit zur Spracherkennung. <http://kaldi-asr.org/> [137]

tes Modul zur Erzeugung simpler GUIs. Ein *Mock-Up* der GUI ist in Abbildung 30 dargestellt.

The mock-up shows a GUI window with a black border. It contains three input fields at the top: the first contains 'bild1.jpg', the second contains 'audio1.wav', and the third is a text input field with the placeholder 'Bitte geben Sie einen Text ein...'. To the right of each input field is a blue button with white text: 'Bild auswählen', 'Audio-Datei auswählen', and 'Text- und Bildanalyse durchführen' respectively. Below the input fields are two more blue buttons with white text: 'Audio- und Bildanalyse durchführen' on the left and 'Text- und Bildanalyse durchführen' on the right. At the bottom left, the text 'Result: Anger 0.87' is displayed.

Abb. 30: *Mock-Up* der GUI für Konzept 2

Die GUI besteht aus zwei Buttons, mit denen jeweils ein Bild und eine Audio-Datei ausgewählt werden können. Die Eingabe eines Textes wird über ein Eingabefeld ist getätigt. Es existieren zwei weitere Buttons, die die jeweilige Analyse durchführen. Einer dieser Buttons führt eine Analyse für eine Audio- und eine Bild-Datei aus. Der andere Button führt eine Analyse für eine Bild-Datei und die Text-Eingabe aus. Das Ergebnis wird in dem unteren Bereich des Fensters ausgegeben.

Auswertung Bild und Text

Die einzelnen Komponenten sollen die eingegeben Daten jeweils für sich verarbeiten. Wie auch im ersten Konzept müssen die einzelnen Ergebnisse zu einem Ergebnis zusammengefügt werden. Die Erkennung durch die Eingabeströme Bild und Text wird ähnlich ausgewertet wie im ersten Konzept, da bei der Erkennung durch Text neben vier Emotionen auch drei Töne ausgegeben werden können, wie auch in Kapitel 5.2.3 erwähnt. Falls keine Emotion erkannt werden kann, muss dies als Hinweis ausgegeben werden. Wenn nur eine Erkennung eine Emotion ermittelt, soll diese ausgegeben werden. Falls beide Erkennungen dieselbe Emotion erkannt

haben (Wut, Trauer, Angst oder Freude), wird diese ausgegeben. Ansonsten werden beide erkannten Emotionen ausgegeben.

Auswertung Bild und Audio

Bei der Erkennung der Eingabeströme Bild und Audio ist es jedoch ungewiss, inwiefern die Emotionen ausgewertet werden müssen. DELTA bietet keine Übersicht darüber, welche Emotionen erkannt werden können. Aus diesem Grund müssen im Zweifelsfall beide Emotionen ausgegeben werden.

Dieser Ansatz sollte eine hohe Genauigkeit in der Erkennung von Emotionen ermöglichen. Des Weiteren besteht hier die Möglichkeit die einzelnen Komponenten jederzeit durch andere auszutauschen. Falls ein besseres Programm oder ein besserer Dienst gefunden wurde, oder auch falls eine selbst-implementierte Lösung hinzugefügt werden soll, können diese jederzeit ausgetauscht werden. Hierbei wird zu jeder Zeit die Funktionalität beibehalten.

Ebenfalls können zusätzliche Eingabeströme hinzugefügt werden. Außerdem können das geschätzte Alter und Geschlecht problemlos als zusätzliche Komponenten genutzt werden. Diesbezüglich müsste nur die Auswertung der Komponenten auf die neuen, zusätzlichen Komponenten angepasst werden.

Im nächsten Kapitel wird die Implementierung der vorgestellten Konzepte und die Probleme währenddessen erläutert. Dabei wird zuerst das erste Konzept und anschließend das zweite Konzept implementiert.

7. Prototypische Implementierung

In diesem Kapitel wird auf die Implementierung der Konzepte für die Anwendung eingegangen. Hierbei werden der Einsatz, die Struktur, die Klassen und die Benutzeroberfläche der Anwendung genauer erläutert. Zum Schluss werden die aufgetretenen Probleme und die Testphase beschrieben.

7.1. Implementierung des ersten Ansatzes

Bereits bei der Installation des *Open-Source*-Projekts "Multimodal Emotion Recognition" traten Probleme auf. Mehrere *Python*-Module aus der "requirements.txt"-Datei konnten nicht installiert werden. Beispielsweise sollte "cv2" installiert werden. "cv2" ist jedoch kein eigenes *Python*-Modul, sondern ein Teil von OpenCV und gehört somit zu dem *Python*-Modul "opencv-python".

Nachdem die benötigten Module installiert werden konnten, wies der Quellcode beim Kompilieren diverse Fehler auf. Hierbei stellte sich heraus, dass mehrere *Python*-Module, die installiert werden müssen, nicht in der "requirements.txt"-Datei angegeben wurden.

Schließlich konnte die Anwendung gestartet werden. Sobald jedoch ein Text eingegeben wurde, um diesen anschließend zu analysieren, trat eine weitere Fehlermeldung auf. Der Fehler hierbei war, dass bestimmte Imports aus anderen Dateien nicht durchgeführt werden konnten. Die Suche nach einer Lösung brachte allerdings keinen Erfolg. Den Fragen in GitHub nach zu urteilen, scheint dies ein häufiges Problem zu sein.

Aus diesem Grund wurde der **Ansatz verworfen**. Dennoch kann gehofft werden, dass die Fehler beseitigt oder andere Lösungen angeboten werden. Sollte dies der Fall sein, kann ein weiterer Versuch der Implementierung erfolgen.

7.2. Implementierung des zweiten Ansatzes

In diesem Kapitel wird die Implementierung des zweiten Ansatzes erläutert. Hierzu werden die einzelnen Schritte, die Nutzung, die Struktur, die Klassen, die Benutzeroberfläche, die aufgetretenen Probleme und die Vorgehensweise beim Testen der Anwendung genauer beschrieben.

Aus Zeitgründen wurde DELTA vorerst nicht installiert und integriert. Die Anwendung wurde dennoch in *Python* implementiert und die beiden Dienste von IBM Watson und Microsoft Azure wurden genutzt.

7.2.1. Einsatz der Anwendung

Aus zeitlichen Gründen wurde auf eine Implementierung auf ROS-E verzichtet. Dennoch ist die **Anwendung so ausgelegt**, dass sie theoretisch **auf ROS-E genutzt werden kann**. Dabei wurden sowohl das Betriebssystem, als auch die Hardware- und Software-Komponenten von ROS-E beachtet.

Die Anwendung wurde in der Programmiersprache *Python* implementiert. Deshalb kann sie auf ROS-E installiert werden. Hierzu müssen nur die *Python*-Module aus der Datei "requirements.txt" installiert werden.

Zum Testen der Anwendung wurde diese vorerst mit **PyInstaller** bereitgestellt. PyInstaller macht es möglich aus einer *Python*-Anwendung eine einzelne ausführbare Datei zu erstellen. Jedoch erstellt PyInstaller eine ausführbare Datei nur für das Betriebssystem, auf dem es installiert worden ist. Aus diesem Grund wurden ausführbare Dateien für die Betriebssysteme MacOS¹ und Windows² erstellt. Diese Dateien sowie die

-
- 1 MacOS ist das Betriebssystem der Computer der Firma Apple. (<https://www.apple.com/de/macos/>)
 - 2 Windows ist das Betriebssystem der Firma Microsoft. (<https://www.microsoft.com/de-de/windows>)

zugehörige Anwenderdokumentation befinden sich auf der beigelegten CD-ROM (Anhang A). [138]

7.2.2. Struktur der Anwendung

Für die Anwendung wurden eine Haupt-Datei ("main.py"), zwei Objekt-Klassen ("analyzed_face.py" und "analyzed_text.py"), eine Übersicht der genutzten *Python*-Module ("requirements.txt"), eine Datei mit verschiedenen Methoden ("util.py") und eine Datei, welche von PyCharm¹ automatisch erstellt wird ("main.spec"), erzeugt. Die Dateien sind in Abbildung 31 aufgelistet.

Abb. 31: Übersicht der Dateien der Anwendung

Die Datei "requirements.txt" enthält die genutzten *Python*-Module. Diese dient der einfachen und schnellen Installation benötigter Module für die Anwendung. Mit folgendem Befehl werden alle Module installiert:

```
pip install -r requirements.txt
```

Die Klassen und die Haupt-Datei werden in den folgenden Kapiteln erläutert.

1 PyCharm ist eine Entwicklungsumgebung der Firma JetBrains für Projekte in der Programmiersprache *Python*. (<https://www.jetbrains.com/de-de/pycharm/>)

7.2.3. Klassen

In der Anwendung wurden zwei Objekt-Klassen erzeugt, um jeweils ein analysiertes Gesicht und einen analysierten Text darzustellen. Zu den Klassen gehören:

- analyzed_face.py
- analyzed_text.py

Des Weiteren wurde eine statische Klasse erzeugt, in welcher allgemeine Methoden implementiert werden können:

- util.py

Auf diese Klassen wird im Folgenden genauer eingegangen.

analyzed_face.py

Objekte dieser Klasse werden "**AnalyzedFace**" bezeichnet. Diese Klasse enthält alle benötigten Informationen über ein analysiertes Gesicht. Zu diesen Informationen gehören:

- face_api_url
- face_api_key
- img_url
- emotions
- age
- gender

Die Variable "face_api_url" enthält die URL für den Dienst zur Emotionserkennung von Microsoft Azure. In der Variable "face_api_key" ist der persönliche Schlüssel zur Nutzung dieses Dienstes gespeichert. Die Variable "img_url" enthält die URL zu dem Bild, welches analysiert werden soll. In der Variable "emotions" (Listing 1) sind die erkannten Emotionen des analysierten Bildes gespeichert. Diese stellt sowohl die erkannten Emotionen dar, als auch die Wahrscheinlichkeit mit der die Emotionen erkannt

wurden. Diese Variable ist vom Typ "dict", was bedeutet, dass sie aus mehreren Werten besteht, die wiederum anderen Werten zugeordnet sind. Somit können alle Emotionen und deren Wahrscheinlichkeiten in einer Variable dargestellt werden. Die Variable "age" enthält das geschätzte Alter der Person, dessen Gesicht analysiert wurde. Die Variable "gender" speichert das Geschlecht der Person, dessen Gesicht analysiert wurde.

```
self.emotions = {"anger": 0,
                  "contempt": 0,
                  "disgust": 0,
                  "fear": 0,
                  "happiness": 0,
                  "neutral": 0,
                  "sadness": 0,
                  "surprise": 0}
```

Listing 1: Variable "emotions" in der Klasse "analyzed_face.py"

Das Objekt enthält einen Konstruktor und eine Methode. Der Konstruktor initialisiert alle Variablen, bis auf "face_api_url" und "face_api_key". Diese haben einen festen Wert und werden nicht geändert. Die URL des Bildes wird in den Konstruktor gegeben. Dieser ruft im letzten Schritt die Methode "analyze_face" auf, in welche die URL weitergegeben wird.

Die Methode "analyze_face" führt eine Emotionserkennung mit dem Dienst von Microsoft Azure durch. Hierzu wird eine *POST-Anfrage* an die URL mit dem entsprechenden Schlüssel und der URL des zu analysierenden Bildes gesendet. Damit wird ein JSON-String zurückgegeben, aus dem die Informationen zu den Emotionen, zum Alter und dem Geschlecht extrahiert und in die zugehörigen Variablen geschrieben werden.

analyzed_text.py

Die Objekte dieser Klasse werden "**AnalyzedText**" bezeichnet. Diese Klasse enthält alle benötigten Informationen über ein analysiertes Gesicht. Zu diesen Informationen gehören:

- tone_api_url
- tone_api_key
- text

- emotions

Die Variable "tone_api_url" enthält die URL für den Dienst zur Emotionserkennung von IBM Watson. In der Variable "tone_api_key" ist der persönliche Schlüssel zur Nutzung dieses Dienstes gespeichert. Die Variable "text" enthält den zu analysierenden Text.

```
self.emotions = {"anger": 0,
                 "fear": 0,
                 "joy": 0,
                 "sadness": 0,
                 "analytical": 0,
                 "confident": 0,
                 "tentative": 0}
```

Listing 2: Variable "emotions" in der Klasse "analyzed_text.py"

In der Variable "emotions" (Listing 2) sind die erkannten Emotionen bzw. Töne des analysierten Textes gespeichert. Diese stellt sowohl die erkannten Emotionen dar, als auch die Wahrscheinlichkeit mit der die Emotionen erkannt wurden. Diese Variable ist vom Typ "dict".

Das Objekt enthält einen Konstruktor und eine Methode. Der Konstruktor initialisiert alle Variablen, bis auf "tone_api_url" und "tone_api_key". Diese haben einen festen Wert und werden nicht geändert. Der zu analysierende Text wird in den Konstruktor gegeben. Dieser ruft im letzten Schritt die Methode "analyze_text" auf, in welche der Text weitergegeben wird.

Die Methode "analyze_text" führt eine Emotionserkennung mit dem Dienst von IBM Watson durch. Hierzu wird eine *POST-Anfrage* mit dem Authentifikator von IBM Watson gesendet. Zum Authentifizieren wird die URL und der entsprechende Schlüssel genutzt. In der *POST-Anfrage* wird der zu analysierende Text mitgegeben, wodurch ein JSON-String zurückgegeben wird. Aus diesem werden die Informationen zu den Emotionen extrahiert und in die Variable geschrieben.

util.py

In der Datei "util.py" werden die Ergebnisse der beiden Dienste miteinander vereint. Dafür ist die Methode "analyze_emotion" zuständig. Diese nimmt ein Objekt der Klasse "AnalyzedFace" und ein Objekt der Klasse "AnalyzedText" an, um sie verarbeiten zu können. Hierbei werden zuerst die

Emotionen mit den höchsten Werten aus den jeweiligen Objekten bestimmt. Mit diesen Emotionen wird das endgültige Ergebnis bestimmt. Dieses wird auf die Art und Weise bestimmt, wie es im Konzept in Kapitel 6.3 beschrieben worden ist. In Listing 3 ist dieser Abschnitt dargestellt.

```
# check scenarios and return string with recognized emotion(s) and its value(s)
if highest_face_emotion_value == 0 and highest_text_emotion_value == 0:
    # both highest values are 0, so no emotion was found in either text oder face recognition
    return "no emotion found"
elif highest_text_emotion_value == 0 and highest_face_emotion_value > highest_text_emotion_value:
    # no emotion was found in text recognition
    return recognized_face_emotion + ": " + str(highest_face_emotion_value)
elif highest_face_emotion_value == 0 and highest_text_emotion_value > highest_face_emotion_value:
    # no emotion was found in face recognition
    return recognized_text_emotion + ": " + str(highest_text_emotion_value)
elif recognized_text_emotion == recognized_face_emotion:
    # both have recognized the same emotion, make average of values
    return recognized_text_emotion + ": " + str(mean([highest_text_emotion_value, highest_face_emotion_value]))
else:
    return recognized_face_emotion + ": " + str(highest_face_emotion_value) + "\n" + recognized_text_emotion + ": " + str(highest_text_emotion_value)
```

Listing 3: Bestimmung der Emotion anhand bereits bestimmter Emotionen und deren Wahrscheinlichkeiten der Emotionserkennungen von IBM Watson und Microsoft Azure

Falls keine Emotion erkannt wird, wird dies als Hinweis ausgegeben. Wenn nur eine Emotion erkannt wurde, soll diese ausgegeben werden. Falls die erkannten Emotionen dieselben sind (Wut, Trauer, Angst oder Freude), wird diese ausgegeben. Die Wahrscheinlichkeit wird hierbei durch den Mittelwert der beiden Erkennungen bestimmt. Falls die Emotionen nicht übereinstimmen, werden beide erkannten Emotionen ausgegeben.

7.2.4. Benutzeroberfläche

Die **Haupt-Datei erzeugt die GUI** der Anwendung. Bei dem Design der Benutzeroberfläche wurde sich grob an das Konzept gehalten. Die GUI ist in Abbildung 32 dargestellt.

Es existieren zwei Eingabefelder. Ein Eingabefeld für die Eingabe des zu analysierenden Textes, das Zweite für die Eingabe der URL des zu analysierenden Bildes.

Des Weiteren existieren drei Buttons. Ein Button beendet die Anwendung.

Abb. 32: GUI der Anwendung

Ein zweiter Button analysiert die eingegebenen Daten. Mit einem Klick auf diesen Button werden je ein Objekt der Klassen "AnalyzedFace" und "AnalyzedText" aus den eingegebenen Daten erzeugt. Diese Objekte werden in die Methode "analyze_emotion" aus der "util.py"-Datei gegeben.

Der letzte Button analysiert Beispiel-Daten. Diese Beispiel-Daten sind in der Datei hinterlegt. Sie können über den Button "Beispieldaten analysieren" analysiert werden. Die Beispiel-Daten bestehen aus einem Text und einer Bild-URL. Weitere Beispiel-Daten wurden im Quellcode auskommentiert. Die Beispiel-Daten lassen sich in der Datei "Beispieldaten.txt" auf der beigelegten CD-ROM (Anhang A) finden. Mit einem Klick auf diesen Button werden je ein Objekt der Klassen "AnalyzedFace" und "AnalyzedText" aus

Abb. 33: GUI der Anwendung nach der Analyse der Beispiel-Daten

den Beispiel-Daten erzeugt. Diese Objekte werden in die Methode "analyze_emotion" aus der "util.py"-Datei gegeben.

Des Weiteren existiert ein Textfeld, in welchem die Ergebnisse aus der Methode "analyze_emotion" ausgegeben werden, wie in Abbildung 33 zu erkennen ist.

Im Gegensatz zum Konzept wurden das Eingabefeld für die Eingabe einer Ton-Datei und der Button für die Analyse einer Bild- und Ton-Datei aufgrund der fehlenden, tonbasierten Emotionserkennung nicht hinzugefügt.

Die **GUI wurde nicht mit Tkinter aufgesetzt**, sondern mit PyQt5. Der Grund hierfür war, dass PyInstaller nicht mit Tkinter kompatibel ist. Die ausführbare Datei konnte nicht auf die Tkinter-Bibliothek zugreifen. PyQt5 ist ein modernes Toolkit zur Erzeugung von Benutzeroberflächen.

7.2.5. Aufgetretene Probleme

Im Laufe der Implementierung der Anwendung sind verschiedene Probleme aufgetreten, die sich meist mittels Recherche oder einer Alternative lösen ließen. Jedoch werden auch Probleme aufgezeigt, für die die Anwendung eine bessere Lösung benötigt.

Die ersten Probleme traten bei der Emotionserkennung von Microsoft Azure auf. Um den Dienst nutzen zu können, muss ein Konto mit einer **privaten E-Mail-Adresse** erstellt werden, bei dem eine **Zahlungsart** angegeben werden musste. Auch wenn keine Zahlungen getätigt werden, musste diese angegeben werden.

Außerdem sind die **URLs** und die **Schlüssel** der Dienste von Microsoft Azure und IBM Watson **im Quellcode als Konstante definiert**. Jedoch sollten diese korrekterweise in einer Datenbank gespeichert und verschlüsselt sein.

Ein weiteres Problem der Emotionserkennung von Microsoft Azure wurde

während der Testphase festgestellt. Bei der Auswahl eines Bildes, auf dem ein Junge abgebildet ist, der die Emotion Angst zeigen sollte, hat die **Emotionserkennung jedoch weder ein Gesicht noch eine Emotion** erkennen können. Eventuell kann es daran liegen, dass der Junge die Hände auf sein Gesicht gelegt hat. (siehe Abbildung 34).

Abb. 34: Bild eines Jungen, der einen ängstlichen Gesichtsausdruck macht. [140]

Das letzte Problem mit der Emotionserkennung von Microsoft Azure ist, dass die **Bilder von einer URL abgerufen** werden müssen. Dies könnte gelöst werden, indem eine Datenbank eingerichtet wird, in der die Bilder kurzzeitig gespeichert werden können.

Weitere Probleme sind mit **PyInstaller** aufgetreten. Für PyInstaller muss eine bestimmte *Python*-Version installiert sein. PyInstaller verweist darauf, dass die *Python*-Versionen 3.5 bis 3.7 unterstützt werden. Hierbei ist es wichtig zu wissen, dass keine höhere Version unterstützt wird, zum Beispiel auch 3.7.1 nicht.

Des Weiteren konnte die GUI nicht mit **Tkinter** aufgesetzt werden, da hier Probleme auftreten. Beim Erzeugen einer ausführbaren Datei wird die Fehlermeldung angezeigt, dass die Module von Tkinter nicht vorhanden sind, obwohl diese standardmäßig in *Python* integriert sind.

7.2.6. Test

In diesem Abschnitt wird die Vorgehensweise der Testphase der Anwendung beschrieben. Die Testphase fiel aufgrund der simplen Anwendung knapp aus. Das Testen wurde nach dem Prinzip **Trial-and-Error** durchgeführt. Es wurden immer Teile der Anwendung fertiggestellt, um sie schritt-

weise testen zu können. Hierfür wurden zuerst die Objekt-Klassen erzeugt und mithilfe von Test-Daten deren Funktionalität getestet. Anschließend wurde die GUI erzeugt und getestet. Im Anschluss daran wurden die Funktionen für die Buttons hinterlegt und getestet. Zum Schluss wurde die gesamte Anwendung mit der Eingabe von verschiedenen Daten überprüft.

8. Ausblick

Die Ergebnisse dieser Masterarbeit können als Grundlage für weitere Projekte und Arbeiten fungieren. Aus diesem Grund werden in diesem Kapitel Ideen für die Anwendung gewonnener Erkenntnisse aufgeführt.

8.1. Implementierung auf ROS-E

Zur Nutzung der Anwendung für ihren vorgesehenen Zweck kann diese auf dem Roboter ROS-E implementiert werden. Wie auch schon in Kapitel 7.2.1 erwähnt, ist die Anwendung bereits darauf ausgelegt.

Hierzu muss sichergestellt sein, dass die korrekte *Python*-Version installiert und eine Internetverbindung vorhanden ist. Es müssen alle *Python*-Module aus der Datei "requirements.txt" installiert werden. Mit diesen Schritten ist es bereits möglich die Anwendung zu nutzen. Dennoch wäre es plausibel, wenn die Herkunft der Daten angepasst werden würde. Statt einer manuellen Eingabe der Daten, können die Daten aus einer Datenbank bezogen werden.

Zur Verbesserung kann außerdem eine Web-Anwendung implementiert werden, welche auf die Hardware von ROS-E ausgelegt wäre.

Um die vollständige Funktionalität auf ROS-E zu gewährleisten, sollte auch die Idee in Kapitel 8.2 realisiert werden. Das Nutzen der Sensoren von ROS-E würde die Funktionalität der Anwendung vervollständigen.

8.2. Nutzung von Live-Daten

Die volle Funktionalität der Anwendung besteht erst, wenn die Daten nicht mehr manuell eingegeben werden müssen oder aus einer Datenbank stammen, sondern direkt von den Sensoren in die Emotionserkennung gegeben werden können. Hier-

bei ist zu beachten, dass die Emotionserkennung durch den Eingabestrom Bild keine Videos verarbeiten kann. Aus diesem Grund muss darauf geachtet werden, dass Bilder verwendet werden. Des Weiteren müssen die Eingänge der GUI mit denen der Sensoren ausgetauscht werden.

8.3. Nutzung von Video-Daten

Die Nutzung von Video-Dateien anstelle von Bild-Dateien kann eine weitere Verbesserung der Funktionalität darstellen. Der Vorteil von Video-Dateien besteht darin, dass mehr Informationen genutzt werden können.

Die Anwendung kann dahingehend verändert werden, dass mehrere Bilder in gleichen Abständen aus dem Video entnommen und zur Analyse genutzt werden. Dies hätte den Vorteil, dass das Gesicht nicht in einer bestimmten Situation analysiert wird, sondern über einen bestimmten Zeitraum. Somit können kurzfristigere Emotionen eliminiert und die wichtigen Emotionen erkannt werden.

Des Weiteren können die Video-Dateien auch in einem KNN genutzt werden, wie es in Abschnitt 8.6 beschrieben wird.

8.4. Erweiterung der Anwendung um den Eingabestrom Ton

Durch den Eingabestrom Ton kann die Anwendung um eine weitere Emotionserkennung erweitert werden, wie auch schon in Kapitel 6.3 erläutert wurde. Die Anwendung kann beispielsweise um einen zweiten Button erweitert werden, welcher eine Emotionserkennung durch den aktiven Eingabestrom Ton und den passiven Eingabestrom Bild durchführt. Hierfür kann das Projekt DELTA integriert werden. Des Weiteren muss eine Objekt-Klasse für den analysierten Ton erzeugt werden. Außerdem muss eine Methode hinzugefügt werden, die die Ergebnisse der beiden Emotionserkennungen auswertet, um daraus eine Emotion zu bestimmen.

Falls die Anwendung mit dem Eingabestrom Ton auf ROS-E erweitert werden soll, muss beachtet werden, dass ROS-E zunächst Ton-Dateien aufnehmen muss. Diese können somit als Eingabe für die Emotionserkennung genutzt werden.

8.5. Einbeziehen zusätzlicher Komponenten

Das Einbeziehen von Alter und Geschlecht können zusätzliche Informationen liefern. Die Komponenten können zwei weitere passive Komponenten darstellen. Alter und Geschlecht werden bereits vom Dienst von Microsoft Azure geschätzt. Sie sind in der Objekt-Klasse des analysierten Bildes abrufbar und können direkt genutzt werden.

Studien zeigten, dass diese Komponenten einen Einfluss auf Emotionen haben können. Jedoch wurden wissenschaftliche Arbeiten dazu nur im Bereich der tonbasierten Emotionserkennung verfasst.

8.6. Nutzung eines KNNs

Die wichtigste Weiterführung der Arbeit besteht darin, eine hybride Emotionserkennung auf Grundlage eines oder mehrerer KNNs zu entwickeln. Hierbei können eigene Eingänge, Ausgänge und Trainingsdaten definiert werden. Die Eingänge beziehen sich hierbei auf die genutzten Eingabeströme.

Die Ausgänge beziehen sich auf die erkannten Emotionen. Somit können die Emotionen aufeinander abgestimmt werden, sodass bei der Nutzung von mehreren KNNs jedes KNN dieselben Ausgänge hat. Das bietet den Vorteil, dass diese genau eine Emotion erkennen können und nicht miteinander in Zusammenhang gebracht werden müssen.

Die Trainingsdaten können je nach Einsatzszenario und gewünschter Genauigkeit ausgewählt werden. Soll die Emotionserkennung nur bei Helligkeit funktionieren? Steht der Roboter in einem großen Raum mit mehreren Menschen und gewisser

Lautstärke? Aufgrund solcher Fragen können die Trainingsdaten ausgewählt werden. Die Trainingsdaten können für Bild-Daten somit mehrere Personen, verschiedene Helligkeiten, Kamerawinkel oder Situationen enthalten. Für Ton-Daten können beispielsweise verschiedene zusätzliche Umgebungsgeräusche enthalten sein.

Mit diesen Möglichkeiten der Konfiguration kann die Anwendung perfekt auf die Einsatzszenarien angepasst werden. Des Weiteren wird sich mit dieser Implementierung nicht von anderen Firmen oder Programmen abhängig gemacht.

9. Zusammenfassung und Fazit

In dieser Masterarbeit wurde eine Anwendung zur hybriden Emotionserkennung entwickelt. Das Ziel war es, einen hybriden Algorithmus zur Emotionserkennung auf Grundlage von wissenschaftlichen Arbeiten und *Open-Source*-Software zu entwickeln.

Zu Beginn dieser Arbeit wurde das Thema konkretisiert und eingegrenzt, indem Einsatzszenarien beschrieben, die Zielgruppe anhand von Personae aufgestellt und Forschungsfragen verfasst wurden.

Anschließend wurde ein Einblick in die Grundlagen sozialer Aspekte gegeben. Hierzu wurden Emotionen analysiert, erläutert wie sich diese zeigen, voneinander unterscheiden und in welchen Modellen diese dargestellt werden können. Des Weiteren wurde auf die aktuelle Kommunikation zwischen Menschen und Roboter eingegangen und gezeigt, wie sie sich durch Emotionserkennung verbessern kann.

Im Anschluss wurden die technischen Grundlagen aufgezeigt. Hierbei wurde der Roboter ROS-E vorgestellt sowie verschiedene Lizenzen für *Open-Source*-Software und die benötigten Teilgebiete der künstliche Intelligenz erläutert.

Daraufhin wurde der Stand der Technik untersucht, indem verschiedene Dienste, *Open-Source*-Software und verschiedene Implementierungen von Emotionserkennungen über verschiedene Eingabeströme vorgestellt wurden.

Basierend auf den Grundlagen wurden zwei Konzepte entwickelt, die darauf aufbauend die Implementierung der Konzepte aufzuzeigen. Die Implementierung der Anwendung befasst sich mit dem Einsatz, der Struktur, den Klassen und der Benutzeroberfläche der Anwendung. Des Weiteren wurden die aufgetretenen Probleme und die Testphase beschrieben.

Zum Schluss wurden Ideen zur Weiterentwicklung der Anwendung und gewonnener Erkenntnisse aufgezeigt.

Die in Kapitel 2.3 gestellten Forschungsfragen konnten im Rahmen dieser Thesis beantwortet werden und im Folgenden durch ihre Kernaussagen zusammengefasst.

In welcher Form können sich Emotionen zeigen?

Emotionen bezeichnen kurzfristige subjektive kognitive Zustände, welche durch innere oder äußere Einflüsse ausgelöst werden können und sich sowohl physisch, als auch psychisch erkennen lassen. Emotionen zeigen sich durch Mimik, Gestik, Stimme oder durch die Veränderung der Herzschläge, Hautleitfähigkeit, Kommunikation zu anderen Menschen oder der eigenen Gedanken. Emotionen stellen jedoch keine Charakterzüge dar.

Wie wirken Roboter auf den Menschen aus heutiger Sicht?

Schon aus heutiger Sicht lässt sich eine gute Kommunikation mit humanoiden Robotern führen. Das häufigste Problem von humanoiden Roboter ist, dass diese kaum auf ihren Gesprächspartner eingehen können. Die Erkennung von Emotionen kann dazu beitragen, dass der Mensch sich besser auf die Kommunikation mit einem humanoiden Roboter einlassen kann.

Wie können Emotionen durch Computer erfasst werden? / Welche Daten können zur Auswertung von Emotionen genutzt werden?

Wissenschaftliche Studien zeigten, dass Bilder, Videos, Audio-Dateien, Texte, das Alter, das Geschlecht, *Eye-Tracking*, die Leitfähigkeit der Haut, die elektrischen Ströme des Gehirns und Herzens und die Aktionsströme der Muskeln zur Auswertung von Emotionen genutzt werden können.

Wichtig hierbei ist, dass nur ein aktiver Eingabestrom und mehrere passive Eingabeströme für eine Emotionserkennung genutzt werden können.

Welche Algorithmen und/oder Dienste gibt es, die die Daten von Emotionen auswerten können?

Es existieren verschiedenste Dienste, *Open-Source*-Software und Ansätze zur Emotionserkennung. Dienste existieren von Microsoft Azure, IBM Watson oder Amazon. *Open-Source*-Software sind beispielsweise FaceApi.js, Sterling Ur Feelings, DELTA oder Multi-modal Emotion Recognition. Die Ansätze zur Emotionserkennung basieren auf Grund der Genauigkeit auf verschiedenen KNNs.

Auffällig ist, dass die meisten Emotionserkennungen Bilder, Videos oder Ton-Dateien verarbeiten.

Welches hybride Konzept, das auf Algorithmen und/oder Open Source Diensten basiert, ermöglicht es einem humanoiden Roboter Emotionen zu erkennen?

Es wurde ein Konzept entwickelt, welches die Emotionserkennung von Microsoft Azure, IBM Watson und DELTA nutzt, um eine hybride Emotionserkennung umsetzen zu können.

Die gewonnenen Erkenntnisse können in weiteren wissenschaftlichen Arbeiten genutzt werden.

Das PDF der Masterarbeit, sowie die ausführbaren Dateien, eine Anwenderdokumentation und Beispiel-Daten befinden sich auf der beigelegten CD-ROM.

"We're fascinated with robots because
they are reflections of ourselves."

- Ken Goldberg

ANHANG A - CD-ROM

Die beiliegende CD-ROM dieser Arbeit hat den folgenden Inhalt:

- **Dokumente:** Dieses Verzeichnis enthält die Masterarbeit als PDF-Dokument, die Anwenderdokumentation des ausführbaren Programms und eine Datei mit Testdaten.
- **Programm:** Dieses Verzeichnis enthält ein weiteres Verzeichnis mit den ausführbaren Dateien der Anwendung.
 - **Mac OS:** Dieses Verzeichnis enthält die ausführbare Datei für Mac OS Systeme.
 - **Windows:** Dieses Verzeichnis enthält die ausführbare Datei für Windows Systeme.

ANHANG B - MIT-LIZENZ

"Copyright <YEAR> <COPYRIGHT HOLDER>

Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE." [76]

ANHANG C - APACHE-2.0-LIZENZ

"Copyright [yyyy] [name of copyright owner]

Licensed under the Apache License, Version 2.0 (the "License");

you may not use this file except in compliance with the License.

You may obtain a copy of the License at

<http://www.apache.org/licenses/LICENSE-2.0>

Unless required by applicable law or agreed to in writing, software

distributed under the License is distributed on an "AS IS" BASIS,

WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.

See the License for the specific language governing permissions and

limitations under the License." [78]

ANHANG D - AGPL-3.0-LIZENZ

"<one line to give the program's name and a brief idea of what it does.>

Copyright (C) <year> <name of author>

This program is free software: you can redistribute it and/or modify it under the terms of the GNU Affero General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

This program is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU Affero General Public License for more details.

You should have received a copy of the GNU Affero General Public License along with this program. If not, see <<https://www.gnu.org/licenses/>>." [79]

GLOSSAR

Affekt	- Affekt beschreibt eine heftige Erregung. [35]
affektiv	- Affektiv bedeutet, dass ein Verhalten von Gefühlen geleitet wird. [36]
autonomes Nervensystem	- Ein autonomes Nervensystem (ANS) entzieht sich weitergehend der willkürlichen Kontrolle und verhält sich deshalb autonom. [32]
Chatbot	- Ein Chatbot ist ein Dialogsystem, bei dem der andere Gesprächspartner automatisiert arbeitet. [38]
Copyright	- Copyright, auch Urheberrecht genannt, ist das Recht über die selbst geschöpften Leistungen allein zu verfügen. [77]
Elektrodermale Aktivität (EDA)	- EDA bezeichnet die Messgröße der Leitfähigkeit der Haut. [62]
Elektroenzephalogramm (EEG)	- Ein EEG ist die Darstellung von elektrischen Strömen des Gehirns. [60]
Elektrokardiogramm (EKG)	- Ein EKG ist die Darstellung von elektrischen Strömen des Herzens. [61]
Elektromyogramm (EMG)	- Ein EEG ist die Darstellung von Aktionsströmen der Muskeln. [63]

Entscheidungsbaum	- Ein Entscheidungsbaum ist eine Darstellung von mehreren Entscheidungen und den daraus resultierenden Ergebnissen.
Extraversion	- Extraversion bezeichnet die Einstellung einer Person, die sich auf äußere Einflüsse konzentriert. [116]
Eye-Tracking	- Eye-Tracking ist die Messung und Aufnahme der Augenbewegungen. [64]
GET-Anfrage	- Mit einer GET-Anfragen können Ressourcen vom Server angefordert werden. [86]
Gradient	- "Ein Gradient misst wie stark sich die Ausgabe einer Funktion ändert, wenn Sie die Eingänge etwas ändern." [102]
Haar-Cascade-Algorithmus	- Der Haar-Cascade-Algorithmus wird im Bereich Machine Learning genutzt, um Objekte in einem Bild oder einem Video zu erkennen. [144]
Hashtag	- Ein Hashtag bezeichnet ein markiertes Schlagwort, welches durch ein Raute-Zeichen gekennzeichnet wird. [145]
Histogramm	- Ein Histogramm ist die Darstellung von Häufigkeitsverteilungen. [146]

integrales Bild	Ein integrales Bild bezeichnet eine effektive Methode, mit welcher die Summen von Pixeln eines Bildes oder einem Abschnitt eines Bildes berechnet werden können. [147]
JavaScript	- JavaScript ist eine webbasierte Programmiersprache, mit der sich das Verhalten einer Webseite programmieren lässt.
Korrelation	- Eine Korrelation bezeichnet einen Zusammenhang zwischen zwei verschiedenen Komponenten. [33]
limbisches System	- Das limbische System ist ein sehr alter Teil des Gehirns. Dieser ist zuständig für die Steuerung von Funktionen, wie Antrieb, Lernen, Gedächtnis, Emotionen, vegetative Regulation der Nahrungsaufnahme, Verdauung und Fortpflanzung. [24]
Mock-Up	- Mock-Ups dienen der ersten Visualisierung einer Webseite.
Neurotizismus	- Neurotizismus bezeichnet die Verfassung einer Person, die Anzeichen von Labilität, Schüchternheit oder Gehemmtheit aufzeigt. [117]
Open Source	- Open Source wird im Zusammenhang mit Software, Programmen oder Ähnlichem verwendet. Open Source bedeutet, dass der Quellcode dieser frei zugänglich ist und, je nach Lizenz, kopiert, genutzt und verändert werden darf. [55]

Perceptual Robotics	- Perceptual Robotics verbindet die Neurobiologie mit Robotik. Dieses Themengebiet befasst sich unter anderem mit Systemen, welche Sensor- und Motorprobleme lösen. [16] [17]
POST-Anfrage	- Mit einer POST-Anfragen können Ressourcen zur Weiterverarbeitung zum Server gesendet werden. [86]
Postulat	- Ein Postulat ist eine Annahme einer Theorie, welche nicht bewiesen oder beweisbar ist. [58]
Python	- Python ist eine Programmiersprache mit einer einfachen Syntax, in welcher Programme kürzer gefasst werden können. [84]
Smart Watch	- Eine Smart Watch ist ein kleiner Computer, der wie eine Uhr am Handgelenk getragen wird. Neben der Zeitanzeige kann eine Smart Watch noch viele weitere Funktionen haben. [57]
Spektrogramm	- Ein Spektrogramm ist die Darstellung eines Signals, wobei der Frequenzgehalt dieses Signals auf die Zeit bezogen wird. [112, S. 87]
Typescript	- Typescript ist eine objektorientierte Programmiersprache, welche Javascript ähnelt und auf ECMA-Script 6 basiert.
Valenz	- Valenz beschreibt eine Wertigkeit, die zeigt, wie sich das Verhalten eines Menschen verändert. [37]

Verarbeitung natürlicher Sprache (NLP) - NLP bezeichnet die maschinelle Verarbeitung natürlicher Sprache. [148]

Wearable - Ein Wearable ist ein System, welches am Körper getragen wird, um die Daten des Nutzers oder der Umwelt zu registrieren und verarbeiten. [56]

ABKÜRZUNGSVERZEICHNIS

API	-	Application Programming Interface (Programmierschnittstelle)
CLM	-	Constrained Local Model (Eingeschränktes Lokales Modell)
CLNF	-	Constrained Local Neural Field (Eingeschränktes Lokales Neuronales Feld)
CNN	-	Convolutional Neural Network (Gefaltetes Neuronales Netzwerk)
EDA	-	Elektrodermale Aktivität
EEG	-	Elektroenzephalogramm
EKG	-	Elektrokardiogramm
EMG	-	Elektromyogramm
FFDP	-	Facial Feature Point Detection (Erkennung von Gesichtsmerkmalen)
GUI	-	Graphic User Interface (Graphische Benutzeroberfläche)
HOG	-	Histogram of Oriented Gradients (Histogramm orientierter Gradienten)
JSON	-	JavaScript Object Notation (JavaScript-Objekt-Notation)
KI	-	Künstliche Intelligenz
KNN	-	Künstliches Neuronales Netz
LSTM	-	Long Short-Term Memory
MIT	-	Massachusetts Institut of Technology (Massachusetts Institut für Technologie)

ML	-	Machine Learning (maschinelles Lernen)
MLP	-	Multilayer Perceptron (Mehrmaliges Perzeptron)
NLP	-	Natural Language Processing ((maschinelle) Verarbeitung natürlicher Sprache)
RNN	-	Recurrent Neural Network
SDK	-	Software Development Kit
SVM	-	Support Vector Machine

ABBILDUNGSVERZEICHNIS

Abbildung 1: Roboter Paul im Elektronikfachhandel Saturn [7]	5
Abbildung 2: Roboter im Elektronikfachhandel Conrad [6]	6
Abbildung 3: Roboter Paula aus einem Pflegeheim in Siegen [8]	6
Abbildung 4: Grundschülerin Lisa Maier [10]	8
Abbildung 5: Malermeister Dirk Kaufmann [11]	9
Abbildung 6: Rentner Martin Müller [12]	10
Abbildung 7: Softwareentwicklerin Laura Schmidt [13]	11
Abbildung 8: Systemtester Christian Schneider [14]	12
Abbildung 9: Das Komponentenmodell der Emotionen [20, S. 168]	17
Abbildung 10: Plutchiks Rad der Emotionen [22]	21
Abbildung 11: Das Circumplexmodell nach Lisa Feldman Barrett und James A. Russell [20, S. 169]	23
Abbildung 12: Therapie Robbe Paro [59]	26
Abbildung 13: Das Trolley-Problem [45]	27
Abbildung 14: Roboter Sophia [19]	28
Abbildung 15: ROS-E mit neutralen Augen [139]	31
Abbildung 16: ROS-E mit einem anderen Ausdruck in den Augen [139]	32
Abbildung 17: Künstliches Neuronales Netz [143]	36

Abbildung 18: Aufbau eines CNNs [131]	38
Abbildung 19: Die drei Grundpfeiler der Emotionserkennung	42
Abbildung 20: Überblick-Seite der multimodalen Emotionserkennung [115]	50
Abbildung 21: Video-Interview-Seite der multimodalen Emotionserkennung [115].....	50
Abbildung 22: Audio-Interview-Seite der multimodalen Emotionserkennung [115].....	51
Abbildung 23: Text-Interview-Seite der multimodalen Emotionserkennung [115]	51
Abbildung 24: Facial Feature Point Detection und die daraus resultierende konvexe Hülle [122]	53
Abbildung 25: Extraktion eines Gesichts mithilfe von FFPD [72].....	56
Abbildung 26: Prozesse der Ende-zu-Ende Emotionserkennung. Abbildung nach [135] ...	60
Abbildung 27: Textbasierte Emotionserkennung mit einem Emotion Embedding Model [27]	63
Abbildung 28: Beispiel zum Finden des emotionalen Wortes eines Satzes. Abbildung nach [27]	65
Abbildung 29: <i>Mock-Up</i> der GUI für Konzept 1	71
Abbildung 30: <i>Mock-Up</i> der GUI für Konzept 2.....	75
Abbildung 31: Übersicht der Dateien der Anwendung	79
Abbildung 32: GUI der Anwendung.....	84
Abbildung 33: GUI der Anwendung nach der Analyse der Beispiel-Daten	84
Abbildung 34: Bild eines Jungen, der einen ängstlichen Gesichtsaus- druck macht. [140]	86

Listing 1: Variable "emotions" in der Klasse "analyzed_face.py"	81
Listing 2: Variable "emotions" in der Klasse "analyzed_text.py"	82
Listing 3: Bestimmung der Emotion anhand bereits bestimmter Emotionen und deren Wahrscheinlichkeiten der Emotionserkennungen von IBM Watson und Microsoft Azure ...	83

TABELLENVERZEICHNIS

Tabelle 1: Emotionen, ihre Funktionen und das dadurch ausgelöste Verhalten. Eigene Tabelle nach [20, S. 169 f.] und [141]	22
Tabelle 2: Genauigkeit der zu erkennenden Emotionen. Tabelle nach [27]	65
Tabelle 3: Beispiel einer Auswertung zur Bestimmung von Emotionen bei unterschied- lichen willkürlichen Ergebnissen.	72

LITERATURVERZEICHNIS

- [1] N. Lücke, "*Studienteilnehmer opfern Menschenleben für einen Roboter*", INGENIEUR.de, Juni 2019. [Online] Verfügbar unter: <https://www.ingenieur.de/technik/fachbereiche/robotik/studienteilnehmer-opfern-menschenleben-fuer-einen-roboter/> [Zugriff am: 04.03.2020]
- [2] J. Vincent, "*AI 'Emotion Recognition' can't be trusted*", THE VERGE, Juli 2019. [Online] Verfügbar unter: <https://www.theverge.com/2019/7/25/8929793/emotion-recognition-analysis-ai-machine-learning-facial-expression-review> [Zugriff am: 04.03.2020]
- [3] IBM, *Tone Analyzer*. [Online] Verfügbar unter: <https://www.ibm.com/watson/services/tone-analyzer/> [Zugriff am: 06.03.2020]
- [4] Amazon, *Emotion Rekognition*. [Online] Verfügbar unter: https://docs.aws.amazon.com/rekognition/latest/dg/API_Emotion.html [Zugriff am: 06.03.2020]
- [5] B. Doerrfeld, "*20+ Emotion Recognition APIs That Will Leave You Impressed, and Concerned*", NORDIC APIS, Dezember 2015. [Online] Verfügbar unter: <https://nordicapis.com/20-emotion-recognition-apis-that-will-leave-you-impressed-and-concerned/> [Zugriff am: 04.03.2020]
- [6] T. Weidemann, "*Roboter im Handel: Conrad stellt Verkaufsroboter vor*", t3n, Oktober 2019. [Online] Verfügbar unter: <https://t3n.de/news/roboter-handel-conrad-stellt-1209740/> [Zugriff am: 02.06.2020]
- [7] G. Rondinella, "*Saturn setzt Roboter 'Paul' für Kundenbetreuung ein*", HORIZONT, November 2016. [Online] Verfügbar unter: <https://www.horizont.net/tech/nachrichten/Pilotprojekt-Saturn-setzt-Roboter-Paul-fuer-Kundenbetreuung-ein-143957> [Zugriff am: 02.06.2020]

- [8] S. Hoock, "*Paula hilft bei Einsamkeit und Langeweile*", WELT, November 2018. [Online] Verfügbar unter: <https://www.welt.de/regionales/nrw/article183584906/Roboter-in-Altenheim-Paula-hilft-bei-Einsamkeit-und-Langeweile.html> [Zugriff am: 02.06.2020]
- [9] Psychomeda, "*Emotion - Lexikon der Psychologie*", Oktober 2016. [Online] Verfügbar unter: <https://www.psychomeda.de/lexikon/emotion.html> [Zugriff am: 24.05.2020]
- [10] Adobe Stock. [Online] Verfügbar unter: <https://stock.adobe.com/de/images/> [Zugriff am: 07.10.2020]
- [11] Adobe Stock. [Online] Verfügbar unter: <https://stock.adobe.com/de/images/positive-craftsman-with-painting-tools-looking-at-camera/355531023> [Zugriff am: 07.10.2020]
- [12] Adobe Stock. [Online] Verfügbar unter: <https://stock.adobe.com/de/images/happy-senior-man/180705627> [Zugriff am: 07.10.2020]
- [13] Adobe Stock. [Online] Verfügbar unter: <https://stock.adobe.com/de/images/porrait-of-female-programmer-in-office/265608067> [Zugriff am: 07.10.2020]
- [14] Adobe Stock. [Online] Verfügbar unter: <https://stock.adobe.com/de/images/35-years-old-man-portrait/207264725> [Zugriff am: 07.10.2020]
- [15] IEEE Xplore, "*Rod Gruppen*". [Online] Verfügbar unter: <https://ieeexplore.ieee.org/author/37283681700> [Zugriff am: 07.10.2020]
- [16] Wikipedia. [Online] Verfügbar unter: https://en.wikipedia.org/wiki/Perceptual_robotics [Zugriff am: 07.10.2020]

- [17] The Laboratory for Perceptual Robotics at Umass Amherst. [Online] Verfügbar unter: <https://www-robotics.cs.umass.edu/> [Zugriff am: 07.10.2020]
- [18] F. Hapgood, "*When Robots Love Among Us*", Discover, Mai 2008. [Online] Verfügbar unter: discovermagazine.com/technology/when-robots-live-among-us [Zugriff am: 07.10.2020]
- [19] derStandard, "'Sophia': Saudi Arabien bürgert Roboter ein", Oktober 2017. [Online] Verfügbar unter: <https://www.derstandard.de/story/2000066761525/sophia-saudi-arabien-buergert-roboter-ein> [Zugriff am: 16.07.2020]
- [20] K. Rothmund, A. Eder, "*Allgemeine Psychologie: Motivation und Emotion*", 1. Auflage, Wiesbaden: VS Verlag für Sozialwissenschaften, 2011.
- [21] T. Kanade, J. F. Cohn, Y. Tian, "*Comprehensive database for facial expression analysis*", 2000 In Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition (pp. 46–53). doi:10.1109/AFGR.2000.840611
- [22] Handelskraft. [Online] Verfügbar unter: <https://www.handelskraft.de/wp-content/uploads/2014/03/Plutchik-Rad-Emotionen.png> [Zugriff am: 07.10.2020]
- [23] R. Schu, "*Big Five Persönlichkeitsfaktoren*", Spektrum.de. [Online] Verfügbar unter: <https://www.spektrum.de/lexikon/psychologie/big-five-persoenlichkeitsfaktoren/2360> [Zugriff am: 16.08.2020]
- [24] DocCheck Flexikon, "*Limbisches System*", April 2016[Online] Verfügbar unter: https://flexikon.doccheck.com/de/Limbisches_System#:~:text=Das%20limbische%20System%20ist%20ein,Nahrungsaufnahme%2C%20Verdauung%20und%20Fortpflanzung%20zugeschrieben. [Zugriff am: 13.10.2020]

- [25] K.K. Imbir, "*Psychoevolutionary Theory of Emotion (Plutchik)*", 2017 In: Zeigler-Hill V., Shackelford T. (eds) *Encyclopedia of Personality and Individual Differences*. Springer, Cham. doi: 10.1007/978-3-319-28099-8_547-1
- [26] H. Karimova, "*The Emotion Wheel: What It Is and How to Use It*", *PositivePsychology*, Juli 2020. [Online] Verfügbar unter: <https://positivepsychology.com/emotion-wheel/> [Zugriff am: 16.08.2020]
- [27] S. Park, B. Bae and Y. Cheong, "Emotion Recognition from Text Stories Using an Emotion Embedding Model," 2020 IEEE International Conference on Big Data and Smart Computing (BigComp), Busan, Korea (South), 2020, pp. 579-583, doi: 10.1109/BigComp48618.2020.00014.
- [28] Microsoft Azure. [Online] Verfügbar unter: <https://azure.microsoft.com/en-us/services/cognitive-services/face/> [Zugriff am: 21.08.2020]
- [29] IBM Watson. [Online] Verfügbar unter: <https://www.ibm.com/de-de/watson> [Zugriff am: 21.08.2020]
- [30] Amazon Rekognition. [Online] Verfügbar unter: https://docs.aws.amazon.com/rekognition/latest/dg/API_Emotion.html [Zugriff am: 14.03.2020]
- [31] M. Götzke, "*App für Asperger-Autisten — Gefühle lesen lernen mit dem Smartphone*", *Deutschlandfunk Kultur*, März 2018 [Online] Verfügbar unter: https://www.deutschlandfunkkultur.de/app-fuer-asperger-autisten-gefuehle-lesen-lernen-mit-dem.1001.de.html?dram:article_id=413486#:~:text=Die%20Neurowissenschaftlerin%20Isabel%20Dziobek%20hat,ganze%20Gesicht%20zu%20sehen%20ist. [Zugriff am: 01.10.2020]

- [32] DocCheck Flexikon, "*Autonomes Nervensystem*". [Online] Verfügbar unter: [https://flexikon.doccheck.com/de/Autonomes Nervensystem](https://flexikon.doccheck.com/de/Autonomes_Nervensystem) [Zugriff am: 07.10.2020]
- [33] Duden, "*Korrelation*". [Online] Verfügbar unter: <https://www.duden.de/rechtschreibung/Korrelation> [Zugriff am: 09.10.2020]
- [34] S. Miller, "*Robert Plutchik, Definer of Animal Emotions*", The New York Sun, Mai 2006. [Online] Verfügbar unter: <https://www.nysun.com/obituaries/robert-plutchik-definer-of-animal-emotions/32264/> [Zugriff am: 22.08.2020]
- [35] Duden, "*Affekt*". [Online] Verfügbar unter: <https://www.duden.de/rechtschreibung/Affekt> [Zugriff am: 09.10.2020]
- [36] Duden, "*affektiv*". [Online] Verfügbar unter: <https://www.duden.de/rechtschreibung/affektiv> [Zugriff am: 09.10.2020]
- [37] Spektrum.de. "*Valenz*". [Online] Verfügbar unter: <https://www.spektrum.de/lexikon/psychologie/valenz/16097> [Zugriff am: 09.10.2020]
- [38] Duden, "*Chatbot*". [Online] Verfügbar unter: <https://www.duden.de/rechtschreibung/Chatbot> [Zugriff am: 09.10.2020]
- [39] J. Wendt, "*Das Wörterbuch 'Mensch-Maschine' wird dicker*", ZEIT, Juli 2014. [Online] Verfügbar unter: <https://www.zeit.de/digital/internet/2014-07/mensch-maschine-kommunikation-gesten-roboter/komplettansicht> [Zugriff am: 21.07.2020]
- [40] Valentin Schröter: Prototypischer Bau eines Tischroboters als Plattform für zukünftige Entwicklungs- und Forschungsprojekte im Bereich des Ambient Assisted Livings und als Alternative zum NAO-Roboter; Bachelorarbeit; Studiengang Telematik

- [41] PARO. [Online] Verfügbar unter: <http://www.parorobots.com/> [Zugriff am: 3.10.2020]
- [42] NLTK. [Online] Verfügbar unter: <http://www.nltk.org/> [Zugriff am: 15.09.2020]
- [43] S. Parsch, "*Wie Roboter dem Menschen ähnlicher werden*", SPIEGEL, März 2018. [Online] Verfügbar unter: <https://www.spiegel.de/wissenschaft/technik/wie-roboter-dem-menschen-aehnlicher-werden-a-1199161.html> [Zugriff am: 07.07.2020]
- [44] Gründerszene, "*Tweet*", Januar 2019. [Online] Verfügbar unter: <https://www.gruenderszene.de/lexikon/begriffe/tweet> [20.10.2020]
- [45] Wikipedia. [Online] Verfügbar unter: https://upload.wikimedia.org/wikipedia/commons/thumb/f/fd/Trolley_Problem.svg/1200px-Trolley_Problem.svg.png [Zugriff am: 10.07.2020]
- [46] kaggle, "*Notes on word embedding algorithms*". [Online] Verfügbar unter: <https://www.kaggle.com/residentmario/notes-on-word-embedding-algorithms> [Zugriff am: 01.10.2020]
- [47] DocCheck Flexikon, "*Trolley-Problem*". [Online] Verfügbar unter: <https://flexikon.doccheck.com/de/Trolley-Problem> [Zugriff am: 10.07.2020]
- [48] R. Bülow, "*Ethik bei autonomen Autos und das Trolley-Problem: Was tut der Weichensteller?*", Heise, Juli 2017. [Online] Verfügbar unter: <https://www.heise.de/newsticker/meldung/Ethik-bei-autonomen-Autos-und-das-Trolley-Problem-Was-tut-der-Weichensteller-3766885.html?seite=2> [Zugriff am: 10.07.2020]

- [49] J. Hacker-Wright, "*Philippa Foot*", Stanford Encyclopedia of Philosophy, August 2018. [Online] Verfügbar unter: <https://plato.stanford.edu/entries/philippa-foot/> [Zugriff am: 02.10.2020]
- [50] Munzinger, "*Karl Engisch*". [Online] Verfügbar unter: <https://www.munzinger.de/search/portrait/Karl+Engisch/0/5399.html> [Zugriff am: 02.10.2020]
- [51] teachtoday, "*Der menschliche Roboter*". [Online] Verfügbar unter: [https://www.teachtoday.de/Informieren/KI im Alltag/2385 Der menschliche Roboter.htm](https://www.teachtoday.de/Informieren/KI_im_Alltag/2385_Der_menschliche_Roboter.htm) [Zugriff am: 16.07.2020]
- [52] J. Ho, "*Wie nah sind Roboter daran, menschlich zu werden?*", Current, Mai 2018. [Online] Verfügbar unter: <https://www.distrelec.de/current/de/kunstliche-intelligenz/wie-nah-sind-roboter-daran-menschlich-zu-werden/> [Zugriff am: 15.07.2020]
- [53] J. Weber, "*Der Roboter als Menschenfreund*", Heise, pp. 140-149, Februar 2006.
- [54] Süddeutsche Zeitung, "*Mein Smartphone weiß, dass ich wütend bin*", März 2019. [Online] Verfügbar unter: <https://www.sueddeutsche.de/digital/smartphone-software-emotionen-simulation-ki-1.4377004-2> [Zugriff am: 16.07.2020]
- [55] Duden, "*Open-Source-Software*". [Online] Verfügbar unter: https://www.duden.de/rechtschreibung/Open_Source_Software [Zugriff am: 05.10.2020]
- [56] Duden, "*Wearable*". [Online] Verfügbar unter: <https://www.duden.de/rechtschreibung/Wearable> [Zugriff am: 05.10.2020]
- [57] Duden, "*Smartwatch*". [Online] Verfügbar unter: <https://www.duden.de/rechtschreibung/Smartwatch> [Zugriff am: 05.10.2020]

- [58] Duden, "*Postulat*". [Online] Verfügbar unter: <https://www.duden.de/rechtschreibung/Postulat> [Zugriff am: 05.10.2020]

- [59] Wohlfahrtswerk. [Online] Verfügbar unter: https://www.wohlfahrtswerk.de/fileadmin/default/_processed_/b/0/csm_Paro_FOC2887_953c6565be.jpg [Zugriff am: 03.10.2020]

- [60] Duden. "*Elektroenzephalogramm*". [Online] Verfügbar unter: <https://www.duden.de/rechtschreibung/Elektroenzephalogramm> [Zugriff am: 05.10.2020]
- [61] Duden. "*Elektrokardiogramm*". [Online] Verfügbar unter: <https://www.duden.de/rechtschreibung/Elektrokardiogramm> [Zugriff am: 05.10.2020]

- [62] Spektrum.de, "*Elektrodermale Aktivität*". [Online] Verfügbar unter: <https://www.spektrum.de/lexikon/psychologie/elektrodermale-aktivitaet/3965> [Zugriff am: 05.10.2020]

- [63] Duden. "*Elektromyogramm*". [Online] Verfügbar unter: <https://www.duden.de/rechtschreibung/Elektromyogramm> [Zugriff am: 05.10.2020]

- [64] Duden. "*Eye-Tracking*". [Online] Verfügbar unter: <https://www.duden.de/rechtschreibung/Eyetracking> [Zugriff am: 05.10.2020]

- [65] W. Zheng, B. Dong and B. Lu, "Multimodal emotion recognition using EEG and eye tracking data," 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Chicago, IL, 2014, pp. 5040-5043, doi: 10.1109/EMBC.2014.6944757.

- [66] H. W. Guo, Y. S. Huang, J. C. Chien and J. S. Shieh, "Short-term analysis of heart rate variability for emotion recognition via a wearable ECG device," 2015 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS), Okinawa, 2015, pp. 262-265, doi: 10.1109/ICIIBMS.2015.7439542.

- [67] A. Greco, C. Marzi, A. Lanata, E. P. Scilingo and N. Vanello, "Combining Electrodermal Activity and Speech Analysis towards a more Accurate Emotion Recognition System," 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 2019, pp. 229-232, doi: 10.1109/EMBC.2019.8857745.
- [68] S. A. Mithbavkar and M. S. Shah, "Recognition of Emotion Through Facial Expressions Using EMG Signal," 2019 International Conference on Nascent Technologies in Engineering (ICNTE), Navi Mumbai, India, 2019, pp. 1-6, doi: 10.1109/ICNTE44896.2019.8945843.
- [69] K. Candra Kirana, S. Wibawanto and H. Wahyu Herwanto, "Facial Emotion Recognition Based on Viola-Jones Algorithm in the Learning Environment," 2018 International Seminar on Application for Technology of Information and Communication, Semarang, 2018, pp. 406-410, doi: 10.1109/ISEMANTIC.2018.8549735.
- [70] Z. Wanli, L. Guoxin and W. Lirong, "Application of Improved Spectral Subtraction Algorithm for Speech Emotion Recognition," 2015 IEEE Fifth International Conference on Big Data and Cloud Computing, Dalian, 2015, pp. 213-216, doi: 10.1109/BDCloud.2015.77.
- [71] J. Umamaheswari and A. Akila, "An Enhanced Human Speech Emotion Recognition Using Hybrid of PRNN and KNN," 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), Faridabad, India, 2019, pp. 177-183, doi: 10.1109/COMITCon.2019.8862221.
- [72] A. Kartali, M. Roglić, M. Barjaktarović, M. Đurić-Jovičić and M. M. Janković, "Real-time Algorithms for Facial Emotion Recognition: A Comparison of Different Approaches," 2018 14th Symposium on Neural Networks and Applications (NEUREL), Belgrade, 2018, pp. 1-4, doi: 10.1109/NEUREL.2018.8587011.

- [73] A. Varghese, J. P. Cherian and J. J. Kizhakkethottam, "Overview on emotion recognition system," 2015 International Conference on Soft-Computing and Networks Security (ICSNS), Coimbatore, 2015, pp. 1-5, doi: 10.1109/ICSNS.2015.7292443.
- [74] D. Verma and D. Mukhopadhyay, "Age driven automatic speech emotion recognition system," 2016 International Conference on Computing, Communication and Automation (ICCCA), Noida, 2016, pp. 1005-1010, doi: 10.1109/ICCA.2016.7813862.
- [75] I. Bisio, A. Delfino, F. Lavagetto, M. Marchese and A. Sciarrone, "Gender-Driven Emotion Recognition Through Speech Signals For Ambient Intelligence Applications," in IEEE Transactions on Emerging Topics in Computing, vol. 1, no. 2, pp. 244-257, Dec. 2013, doi: 10.1109/TETC.2013.2274797.
- [76] Open Source Initiative, "*The MIT License*". [Online] Verfügbar unter: <https://open-source.org/licenses/MIT> [Zugriff am: 27.08.2020]
- [77] Duden. "*Urheberrecht*". [Online] Verfügbar unter: <https://www.duden.de/recht-schreibung/Urheberrecht> [Zugriff am: 27.08.2020]
- [78] The Apache Software Foundation, "*APACHE LICENSE, VERSION 2.0*". [Online] Verfügbar unter: <http://www.apache.org/licenses/LICENSE-2.0> [Zugriff am: 28.08.2020]
- [79] GNU, "*GNU Affero General Public License*". [Online] Verfügbar unter: <https://www.gnu.org/licenses/agpl-3.0.html> [Zugriff am: 01.09.2020]
- [80] GNU, "*Copyleft. Was ist das?*". [Online] Verfügbar unter: <https://www.gnu.org/licenses/copyleft.de.html> [Zugriff am: 02.09.2020]

- [81] GNU, "*Freie Software. Was ist das?*". [Online] Verfügbar unter: <https://www.gnu.org/philosophy/free-sw> [Zugriff am: 02.09.2020]
- [82] AI United. "*7 Arten künstlicher neuronaler Netzwerke in der linguistischen Datenverarbeitung*". [Online] Verfügbar unter: <http://www.ai-united.de/7-arten-kuenstlicher-neuronaler-netzwerke-in-der-linguistischen-datenverarbeitung/> [Zugriff am: 12.09.2020]
- [83] Microsoft Azure. "*Pricing Calculator*". [Online] Verfügbar unter: <https://azure.microsoft.com/de-de/pricing/calculator/> [Zugriff am: 21.08.2020]
- [84] W3Schools, "*Python Introduction*". [Online] Verfügbar unter: https://www.w3schools.com/python/python_intro.asp [Zugriff am: 15.08.2020]
- [85] GitHub, "*face-api.js*". [Online] Verfügbar unter: <https://github.com/justadudewhohacks/face-api.js> [Zugriff am: 03.09.2020]
- [86] Selfhtml, "*HTTP/Anfragemethoden*". [Online] Verfügbar unter: <https://wiki.selfhtml.org/wiki/HTTP/Anfrage-methoden> [Zugriff am: 15.08.2020]
- [87] Microsoft Cognitive Services, "*Face API - v1.0*". [Online] Verfügbar unter: <https://westus.dev.cognitive.microsoft.com/docs/services/563879b61984550e40cbb8d/operations/563879b61984550f30395236> [Zugriff am: 01.10.2020]
- [88] Guilford Press, "*James A. Russell*". [Online] Verfügbar unter: <https://www.guilford.com/author/James-A-Russell> [Zugriff am: 5.10.2020]
- [89] Lisa Feldman Barrett. [Online] Verfügbar unter: <https://lisafeldmanbarrett.com/about/> [Zugriff am: 05.10.2020]

- [90] Node.js. [Online] Verfügbar unter: <https://nodejs.org/de/about/> [Zugriff am: 06.10.2020]
- [91] Springer Gabler Wirtschaftslexikon "maschinelles Lernen", Februar 2018. [Online] Verfügbar unter: <https://wirtschaftslexikon.gabler.de/definition/maschinelles-lernen-38193/version-261619> [Zugriff am: 17.09.2020]
- [92] K. Jaedtke, "*WAS IST MASCHINELLES LERNEN?*", Lernen wie Maschinen, April 2018. [Online] Verfügbar unter: <https://www.lernen-wie-maschinen.ai/ki-pedia/was-ist-maschinelles-lernen/> [Zugriff am: 17.09.2020]
- [93] GitHub, "*STEALING UR FEELINGS*". [Online] Verfügbar unter: <https://github.com/noahlevenson/stealing-ur-feelings> [Zugriff am: 03.09.2020]
- [94] MacArthur Foundation. [Online] Verfügbar unter: <https://www.macfound.org/about/> [Zugriff am: 06.10.2020]
- [95] Ford Foundation. [Online] Verfügbar unter: <https://www.fordfoundation.org/about/about-ford/> [Zugriff am: 06.10.2020]
- [96] Knight Foundation. [Online] Verfügbar unter: <https://knightfoundation.org/about/> [Zugriff am: 06.10.2020]
- [97] Open Society Foundations. [Online] Verfügbar unter: <https://www.opensociety-foundations.org/> [Zugriff am: 06.10.2020]
- [98] Forbes. [Online] Verfügbar unter: <https://www.forbes.com/> [Zugriff am: 06.10.2020]
- [99] CBC. [Online] Verfügbar unter: <https://www.cbc.ca/> [Zugriff am: 06.10.2020]

- [100] GitHub. "*EmoPy*". [Online] Verfügbar unter: <https://github.com/thoughtworksarts/EmoPy> [Zugriff am: 26.08.2020]
- [101] A. Perez, "*EmoPy: A Machine Learning Toolkit For Emotional Expression*", ThoughtWorks ARTS, September 2018. [Online] Verfügbar unter: <https://thoughtworksarts.io/blog/emopy-emotional-expression-toolkit/> [Zugriff am: 26.08.2020]
- [102] AI United "*Recurrent Neural Networks und LSTM*". [Online] Verfügbar unter: <http://www.ai-united.de/recurrent-neural-networks-und-lstm/> [Zugriff am: 02.10.2020]
- [103] ThoughtWorks ARTS. [Online] Verfügbar unter: <https://www.thoughtworks.com/de/about-us> [Zugriff am: 26.08.2020]
- [104] ThoughtWorks ARTS, "*Karen Palmer*". [Online] Verfügbar unter: <https://thoughtworksarts.io/bio/karen-palmer/> [Zugriff am: 26.08.2020]
- [105] S. Luber, N. Litzel, "*Was ist ein Neuronales Netz?*", BIGDATA INSIDER, Februar 2018. [Online] Verfügbar unter: <https://www.bigdata-insider.de/was-ist-ein-neuronales-netz-a-686185/> [Zugriff am: 02.10.2020]
- [106] A. Klofat, "*Wie funktioniert Natural Language Processing in der Praxis? Ein Überblick*", Data Science Blog, Januar 2020. [Online] Verfügbar unter: <https://data-science-blog.com/blog/2020/01/24/wie-funktioniert-natural-language-processing-in-der-praxis-ein-uberblick-dr-aleksandra-klofat/> [Zugriff am: 02.10.2020]
- [107] GitHub, "*DELTA*". [Online] Verfügbar unter: <https://github.com/didi/delta> [Zugriff am: 27.08.2020]
- [108] DELTA, "*Documentation*". [Online] Verfügbar unter: <https://delta-didi.readthedocs.io/en/latest/introduction.html> [Zugriff am: 27.08.2020]

- [109] L. Wuttke, "*Machine Learning vs. Deep Learning: Wo ist der Unterschied?*", datasolut. [Online] Verfügbar unter: <https://datasolut.com/machine-learning-vs-deep-learning/> [Zugriff am: 27.08.2020]
- [110] K. Han, J. Chen, H. Zhang, H. Xu, Y. Peng, Y. Wang, N. Ding, H. Deng, Y. Gao, T. Guo, Y. Zhang, Y. He, B. Ma, Y. Zhou, K. Zhang, C. Liu, Y. Lyu, C. Wang, C. Gong, Y. Wang, W. Zou, H. Song, X. Li, "*DELTA: A DEep learning based Language Technology plAtform*", 2019, arXiv: 1908.01853.
- [111] GitHub. "*Speech Emotion Analyzer*". [Online] Verfügbar unter: <https://github.com/MITESHPUTHRANNEU/Speech-Emotion-Analyzer> [Zugriff am: 28.08.2020]
- [112] M. Werner, "*Digitale Signalverarbeitung mit MATLAB®*", Springer Fachmedien Wiesbaden, pp. 87-112, doi: 10.1007/978-3-658-18647-0
- [113] IBM Tone Analyzer, "*Preisstrukturpläne*". [Online] Verfügbar unter: <https://cloud.ibm.com/catalog/services/tone-analyzer> [Zugriff am: 29.08.2020]
- [114] R. Rojas, "*Theorie der neuronalen Netze*", Springer: Heidelberg, 1993, pp. 149-172, doi: 10.1007/978-3-642-61231-2
- [115] GitHub, "*Multimodal Emotion Recognition*". [Online] Verfügbar unter: <https://github.com/maelfabien/Multimodal-Emotion-Recognition> [Zugriff am: 13.08.2020]
- [116] Duden, "*Extraversion*". [Online] Verfügbar unter: <https://www.duden.de/recht-schreibung/Extraversion> [Zugriff am: 7.10.2020]
- [117] Duden, "*Neurotizismus*". [Online] Verfügbar unter: <https://www.duden.de/recht-schreibung/Neurotizismus> [Zugriff am: 7.10.2020]

- [118] T. Baltrušaitis, A. Zadeh, Y. C. Lim and L. Morency, "OpenFace 2.0: Facial Behavior Analysis Toolkit", *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, Xi'an, 2018, pp. 59-66, doi: 10.1109/FG.2018.00019.
- [119] Matlab. [Online] Verfügbar unter: <https://de.mathworks.com/products/matlab.html> [Zugriff am: 04.10.2020]
- [120] A. Zadeh, T. Baltrušaitis and L. Morency, "Convolutional Experts Constrained Local Model for Facial Landmark Detection," *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Honolulu, HI, 2017, pp. 2051-2059, doi: 10.1109/CVPRW.2017.256.
- [121] T. Baltrušaitis, P. Robinson and L. Morency, "Constrained Local Neural Fields for Robust Facial Landmark Detection in the Wild," *2013 IEEE International Conference on Computer Vision Workshops*, Sydney, NSW, 2013, pp. 354-361, doi: 10.1109/ICCVW.2013.54.
- [122] Petapixel. [Online] Verfügbar unter: https://petapixel.com/assets/uploads/2016/06/facialrecognition_1.jpg [Zugriff am: 17.09.2020]
- [123] S. Mallick, "*Histogram of Oriented Gradients*", Learn OpenCV, Dezember 2016. [Online] Verfügbar unter: <https://www.learnopencv.com/histogram-of-oriented-gradients/> [Zugriff am: 28.09.2020]
- [124] S. Luber, N. Litzel, "*Was ist eine Support Vector Machine?*", BIGDATA INSIDER, November 2019. [Online] Verfügbar unter: <https://www.bigdata-insider.de/was-ist-eine-support-vector-machine-a-880134/> [Zugriff am: 11.10.2020]

- [125] S. Luber, N. Litzel, "*Was ist ein Perzeptron?*", BIGDATA INSIDER, Februar 2019. [Online] Verfügbar unter: <https://www.bigdata-insider.de/was-ist-ein-perzeptron-a-798367/> [Zugriff am: 13.10.2020]

- [126] S. Nayak, "*Understanding AlexNet*", Learn OpenCV, Juni 2018. [Online] Verfügbar unter: <https://www.learnopencv.com/understanding-alexnet/> [Zugriff am: 14.10.2020]

- [127] A. Krizhevsky, I. Sutskever, G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," Advances in Neural Information Processing Systems 25, Curran Associates. Inc. 2012, pp. 1097-1105, <http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>. [Zugriff am: 14.10.2020]

- [128] J. Brownlee, "*A Gentle Introduction to Transfer Learning for Deep Learning*", Machine Learning Mastery, Dezember 2017. [Online] Verfügbar unter: <https://machinelearningmastery.com/transfer-learning-for-deep-learning/> [Zugriff am: 15.10.2020]

- [129] R. Ruizendaal, "*Deep Learning #3: More on CNNs & Handling Overfitting*", towards data science, May 2017. [Online] Verfügbar unter: <https://towardsdatascience.com/deep-learning-3-more-on-cnns-handling-overfitting-2bd5d99abe5d> [Zugriff am: 15.10.2020]

- [130] S. Luber, N. Litzel, "*Was ist ein Convolutional Neural Network?*", BIGDATAT INSIDER, Februar 2019. [Online] Verfügbar unter: <https://www.bigdata-insider.de/was-ist-ein-convolutional-neural-network-a-801246/> [Zugriff am: 15.10.2020]

- [131] Zschech, P., Sager, C., Siebers, P. *et al.* Mit Computer Vision zur automatisierten Qualitätssicherung in der industriellen Fertigung: Eine Fallstudie zur Klassifizierung von Fehlern in Solarzellen mittels Elektrolumineszenz-Bildern. *HMD* (2020). <https://doi.org/10.1365/s40702-020-00641-8>

- [132] J. Brownlee, "*Understand the Impact of Learning Rate on Neural Network Performance*", Machine Learning Mastery, Januar 2019. [Online] Verfügbar unter: <https://machinelearningmastery.com/understand-the-dynamics-of-learning-rate-on-deep-learning-neural-networks/> [Zugriff am: 15.10.2020]

- [133] J. Brownlee, "*Difference Between a Batch and an Epoch in a Neural Network*", Machine Learning Mastery, Juli 2018. [Online] Verfügbar unter: <https://machinelearningmastery.com/difference-between-a-batch-and-an-epoch/> [Zugriff am: 15.10.2020]

- [134] Daniel McDuff, Abdelrahman Mahmoud, Mohammad Mavadati, May Amr, Jay Turcot, and Rana el Kaliouby. 2016. AFFDEX SDK: A Cross-Platform Real-Time Multi-Face Expression Recognition Toolkit. In Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '16). Association for Computing Machinery, New York, NY, USA, 3723–3726. DOI:<https://doi.org/10.1145/2851581.2890247>

- [135] H. Zhao, N. Ye and R. Wang, "A Survey on Automatic Emotion Recognition Using Audio Big Data and Deep Learning Architectures," 2018 IEEE 4th International Conference on Big Data Security on Cloud (BigDataSecurity), IEEE International Conference on High Performance and Smart Computing, (HPSC) and IEEE International Conference on Intelligent Data and Security (IDS), Omaha, NE, 2018, pp. 139-142, doi: 10.1109/BDS/HPSC/IDS18.2018.00039.

- [136] Conda. [Online] Verfügbar unter: <https://docs.conda.io/en/latest/> [Zugriff am: 16.10.2020]

- [137] Kaldi [Online] Verfügbar unter: <http://kaldi-asr.org/> [Zugriff am: 16.10.2020]

- [138] PyInstaller [Online] Verfügbar unter: <http://www.pyinstaller.org/> [Zugriff am: 20.10.2020]

- [139] Fotos von Valentin Schröter
- [140] The Christian Science Monitor. [Online] Verfügbar unter: https://images.csmonitor.com/csm/2014/12/12-17halone_standard_600x400.png?alias=standard_900x600 [Zugriff am: 15.10.2020]
- [141] PositivePsychology. [Online] Verfügbar unter: <https://positivepsychology.com/wp-content/uploads/wheel-of-emotions1.png> [Zugriff am: 16.08.2020]
- [142] M. Siepermann, "Künstliche Intelligenz (KI)", Springer Gabler Wirtschaftslexikon. [Online] Verfügbar unter: <https://wirtschaftslexikon.gabler.de/definition/kuenstliche-intelligenz-ki-40285> [Zugriff am: 16.10.2020]
- [143] AI-United. [Online] Verfügbar unter: <http://www.ai-united.de/wp-content/uploads/2019/01/ann.png> [Zugriff am: 16.10.2020]
- [144] OpenCV Dokumentation. [Online] Verfügbar unter: https://opencv-python-tutroals.readthedocs.io/en/latest/py_tutorials/py_objdetect/py_face_detection/py_face_detection.html [Zugriff am: 16.10.2020]
- [145] "Hashtag", Duden. [Online] Verfügbar unter: <https://www.duden.de/rechtschreibung/Hashtag> [Zugriff am: 16.10.2020]
- [146] "Histogramm", Statistik-Nachhilfe. [Online] Verfügbar unter: <https://www.statistik-nachhilfe.de/ratgeber/statistik/deskriptive-statistik/visualisierung-von-daten/histogramm> [Zugriff am: 16.10.2020]
- [147] "Computer Vision — The Integral Image", Computer Science Source, September 2010. [Online] Verfügbar unter: <https://computersciencesource.wordpress.com/2010/09/03/computer-vision-the-integral-image/> [Zugriff am: 16.10.2020]

- [148] S. Luber, N. Litzel, "Was ist ein Natural Language Processing?", BIGDATAT INSIDER, September 2016. [Online] Verfügbar unter: <https://www.bigdata-insider.de/was-ist-natural-language-processing-a-590102/> [Zugriff am: 16.10.2020]