

**DRAGON: ERFASSUNG VON BEFUNDEN MITTELS
SPRACHERKENNUNG**

Eine Bachelorarbeit von

JOHANNA LANZ

betreut von

DI Dr. techn. Markus Bödenler, BSc

und eingereicht am

Studiengang eHealth

der

FH JOANNEUM Graz

zur teilweisen Erfüllung der Anforderungen

zur Erlangung des akademischen Grades

Bachelor of Science (BSc)

Durch das Ohr der Technik sprechen wir die Sprache des Fortschritts - GPT

Danksagung

Ich möchte diese Gelegenheit nutzen, um mich bei meinem Betreuer DI Dr. techn. Markus Bödenler, BSc herzlich zu bedanken. Seine fachliche Expertise, Geduld und das stetige Engagement waren während der Verfassung meiner Bachelorarbeit von unschätzbarem Wert. Besonders seine fundierten Ratschläge haben maßgeblich zu meinem Erfolg beigetragen. Ebenso gilt mein Dank der IT-Abteilung des Krankenhauses der Barmherzigen Brüder Marschallgasse. Die technische Unterstützung und die bereitgestellten Ressourcen waren essenziell für die Durchführung meiner Forschung.

Ein besonderer Dank geht auch an Herrn Wolfgang Mayer Dipl. FW., der mir nicht nur das Thema zur Verfügung stellte, sondern auch tiefgreifende Einblicke in den klinischen Alltag gewährte.

Darüber hinaus möchte ich allen Teilnehmern, die den Fragebogen ausgefüllt haben, meinen aufrichtigen Dank aussprechen. Ihre wertvollen Beiträge und ehrlichen Antworten haben entscheidend zur Qualität und Aussagekraft meiner Forschung beigetragen.

Abschließend möchte ich meiner Familie meinen tiefsten Dank aussprechen. Ihre unermüdliche Unterstützung und Geduld während meiner gesamten Studienzeit waren eine enorme Stütze für mich.

Ich danke Ihnen allen von ganzem Herzen für Ihre Unterstützung und Ermutigung.

Abstract

This bachelor's thesis investigates the implementation of technology in healthcare, focusing on the "Krankenhaus der Barmherzigen Brüder" in Graz, Austria. The central theme is the deployment of "Dragon Medical SpeechKit". The study covers the evolution of speech recognition technologies, their key applications in healthcare, and a critical examination of their underlying technologies and acceptance among healthcare professionals.

The goal of this research is to evaluate the impact of speech recognition on improving the quality, comprehensiveness, and efficiency of medical documentation. The thesis explores the challenges of integrating these technologies into existing clinical workflows, supported by a literature review and insights gathered from a user acceptance survey among registered nurses and an observational analysis of the introduction of "Dragon Medical SpeechKit" into clinical practice.

Overall, the speech recognition technology "Dragon Medical SpeechKit" was well received. Users reported that this provision increased the speed and accuracy of documentation. The text modules, which lead to shorter documentation times, are particularly appreciated. Despite the positive experiences, there were some challenges, such as technical problems and concerns about the accuracy of word recognition. Potential for improvement was also identified in the training of the technology.

Kurzfassung

Diese Bachelorarbeit untersucht die Implementierung von Technologien im Gesundheitswesen am Beispiel des „Krankenhaus der Barmherzigen Brüder“ in Graz, Österreich. Im Mittelpunkt steht der Einsatz des „Dragon Medical SpeechKit“. Die Studie behandelt die Entwicklung von Spracherkennungstechnologien, ihre wesentlichen Anwendungen im Gesundheitsbereich sowie eine kritische Betrachtung ihrer zugrunde liegenden Technologien und ihrer Akzeptanz bei den Gesundheitsfachkräften.

Ziel dieser Untersuchung ist es, die Auswirkungen der Spracherkennung auf die Verbesserung der Qualität, Vollständigkeit und Effizienz der medizinischen Dokumentation zu bewerten. Die Thesis untersucht die Herausforderungen bei der Integration dieser Technologien in bestehende klinische Arbeitsabläufe, unterstützt durch eine Literaturrecherche und Erkenntnisse aus einer Umfrage zur Benutzerakzeptanz unter Krankenpflegern sowie einer Beobachtungsanalyse zur Einführung des „Dragon Medical SpeechKit“ in den klinischen Alltag.

Insgesamt wurde die Spracherkennungstechnologie „Dragon Medical SpeechKit“ positiv angenommen. Nutzer:innen berichteten, dass durch diese Bereitstellung die Dokumentationsgeschwindigkeit und -genauigkeit gesteigert wurde. Besonders geschätzt werden die Textbausteine, die zu einer verkürzten Dokumentationszeit führen. Trotz der positiven Erfahrungen gab es einige Herausforderungen, wie technische Probleme und Bedenken bei der Genauigkeit der Worterkennung. Weiters wurde Verbesserungspotenzial bei den Schulungen aufgezeigt.

Inhaltsverzeichnis

Danksagung	iii
Abstract	iv
Kurzfassung	v
Abbildungsverzeichnis	viii
Tabellenverzeichnis	ix
1 Einleitung	1
1.1 Ausgangslage	1
1.2 Ziele und Forschungsfragen der Arbeit	2
1.3 Aufbau der Arbeit	2
2 Spracherkennung	4
2.1 Definition und Grundprinzipien	4
2.1.1 Arten von Spracherkennungstechnologien	5
2.1.2 Vokabular	6
2.1.3 Sprachmodelle	6
2.1.4 Umwandlung	7
2.1.5 Qualitätskriterien	8
2.2 Geschichte der Spracherkennung	9
2.3 Branchen	10
2.4 Ansätze	11
2.4.1 Maschinelles Lernen	11
2.4.2 Deep Learning	13
3 Spracherkennung im Gesundheitswesen	16
3.1 Medizinische Befunderhebung	16
3.2 Meilensteine der Vergangenheit	17

3.3	Anwendungsbereiche im Gesundheitswesen	19
3.3.1	Datenbanken	21
3.4	Vorteile und Herausforderungen	22
3.4.1	Datenschutzaspekte	25
3.5	Unternehmen	25
4	Methoden	27
4.1	Literaturrecherche	27
4.2	Interview	29
4.3	Beobachtungsanalyse	29
4.4	Fragebogen	30
5	Resultate	31
5.1	Dragon Medical SpeechKit	31
5.1.1	Einführung in den klinischen Workflow	33
5.2	Datenanalyse Beobachtungsanalyse	34
5.3	Datenanalyse Fragebogen	37
6	Diskussion	47
6.1	Zusammenfassung	47
6.2	Limitierung	48
6.3	Reflexion und Zukunftsausblick	49
	Literaturverzeichnis	51
	A Fragebogen	61
	B Interview Transkript	63
	C Rohdaten Fragebogen	65
	Eidesstattliche Erklärung	72

Abbildungsverzeichnis

2.1	Darstellung der Beziehung von künstlicher Intelligenz, maschinellen Lernens und Deep Learning	12
4.1	Auswahlprozess der Literatur nach den PRISMA Richtlinien.	28
5.1	Philips SpeekMike Premium Touch 3720	36
5.2	Auswertung Fragebogen: Kategorie Benutzerfreundlichkeit	37
5.3	Auswertung Fragebogen: Kategorie Genauigkeit und Vollständigkeit	39
5.4	Auswertung Fragebogen: Kategorie Effizienz	41
5.5	Auswertung Fragebogen: Kategorie Schulungen	43
5.6	Auswertung Fragebogen: Kategorie Zufriedenheit und allgemeine Meinung	44

Tabellenverzeichnis

5.1	Vergleich von Spracherkennungssysteme	33
5.2	Überblick der häufigsten Beobachtungen während der Beobachtungsanalyse	35

Kapitel 1

Einleitung

In einer Ära, in der Datenverarbeitungsmengen rapide zunehmen, ist laut Malik et al. [2021] die effiziente Verarbeitung und Verwaltung dieser Daten zu einem zentralen Anliegen geworden. Die automatisierte Dokumentation, besonders durch den Einsatz von Spracherkennungstechnologien, stellt eine vielversprechende Lösung dar. Diese Technologien ermöglichen eine schnelle und genaue Umwandlung von gesprochener Sprache in Text, was eine Zeitersparnis und Effizienzsteigerung mit sich bringt. Die Relevanz der automatisierten Dokumentation wird laut Kumar [2024] durch ihre vielfältigen Anwendungsbereiche unterstrichen, von der medizinischen Dokumentation bis hin zur Verarbeitung juristischer Texte und weiter bis in die Pädagogik. Zudem tragen Spracherkennungstechnologien laut Bakst et al. [2023] dazu bei, Wissensbarrieren abzubauen und den Zugang zu Informationen zu demokratisieren, indem sie eine barrierefreie Informationsverarbeitung ermöglichen.

1.1 Ausgangslage

Im Krankenhaus der Barmherzigen Brüder Graz soll bis Ende 2024 die klinische Dokumentation der Pflege vom Tippen auf einer Tastatur auf Spracherkennung umgestellt werden. Die Barmherzigen Brüder Graz haben sich für das Programm "Dragon Medical SpeechKit" Nuance [2016], des Unternehmens Nuance Onitilo et al. [2023] entschieden, da das Schwesterprogramm "Dragon Medical One" von Nuance [2024] bereits bei den Ärzt:innen seit 2022 benutzt wird und es keine vergleichbaren Anwendungen für medizinische Spracherkennungssysteme in der Krankenhausumgebung am Markt gibt. Im Zuge dieser Arbeit wird die Einführung in den klinischen Workflow beschrieben, zusätzlich wird die Benutzerakzeptanz des Spracherkennungssystems durch einen Fragebogen und eine Beobachtungsanalyse untersucht.

1.2 Ziele und Forschungsfragen der Arbeit

Diese Arbeit befasst sich mit der Spracherkennung im Allgemeinen und geht daraufhin auf den Einsatz im Gesundheitswesen ein. Dabei werden sowohl die Historie, als auch die technische Umsetzung des Themas beleuchtet. Weiters beschäftigt sich die Arbeit mit einer Befragung von diplomierten Gesundheits- und Krankenpfleger:innen, in der die Bewertung der Akzeptanz von Befunderfassung mit Spracherkennungstechnologien im Krankenhaus der Barmherzigen Brüder Graz beschrieben wird. Weiters soll eine ganzheitliche Bewertung durch eine Beobachtungsanalyse stattfinden, die die Einführung einer Spracherkennungstechnologie wie "Dragon Medical SpeechKit" in einen klinischen Workflow beschreibt. Dabei wird sowohl auf die Vorteile, als auch auf die Nachteile beziehungsweise Herausforderungen eingegangen.

Die Forschungsfragen, die in dieser Bachelorarbeit beantwortet werden sollen, sind:

- Inwiefern beeinflusst die Integration von "Dragon", einer Spracherkennungstechnologie, die Qualität der Befunderfassung im Gesundheitswesen, insbesondere im Hinblick auf Genauigkeit, Vollständigkeit und zeitliche Effizienz und welche Herausforderungen können bei der erfolgreichen Integration in den klinischen Workflow auftreten?
- Wie beeinflusst die Implementierung von "Dragon" die Akzeptanz der Befunderfassung mit Spracherkennungstechnologien bei den Barmherzigen Brüdern Graz, insbesondere unter Berücksichtigung von Aspekten wie Nutzerfreundlichkeit und Anwenderzufriedenheit?

1.3 Aufbau der Arbeit

Die vorliegende Arbeit gliedert sich in sieben Kapitel, die sowohl die theoretischen Grundlagen von Spracherkennung, als auch die praktische Umsetzung der Umstellung der klinischen Dokumentation auf das Spracherkennungssystem "Dragon Medical SpeechKit" beschreiben soll.

Im ersten Kapitel werden der Hintergrund sowie die Ziele und Forschungsfragen beschrieben. Es folgt eine Zusammenfassung der Kapitelinhalte.

Im zweiten Kapitel wird ein Überblick über die Spracherkennung gegeben. Zu Beginn werden die Definition und Grundlagen eingeführt, warum Spracherkennung überhaupt notwendig ist und wie Sprache in unserem Körper entsteht. Im zweiten Punkt wird die Umwandlung von Sprache in Text in sieben Modulen beschrieben und anschließend ein

Überblick über die Entwicklung der Spracherkennung gegeben. Abschließend werden die Branchen beschrieben, in denen Spracherkennung eingesetzt wird. Zusätzlich werden auch die Ansätze, zur Erstellung von Spracherkennungssystemen erklärt.

Das dritte Kapitel befasst sich mit der Spracherkennung speziell im Gesundheitswesen, wobei zunächst geklärt wird, warum die Spracherkennung im Gesundheitswesen das Potenzial hat, eine wichtige Rolle zu spielen. Anschließend werden die Meilensteine der Entwicklung von Spracherkennungssystemen im Gesundheitswesen erläutert. Weiters werden spezielle medizinische Bereiche genannt, in denen Spracherkennungstechnologien zum Einsatz kommen. Daraufhin werden die Vorteile, sowie die Herausforderungen beleuchtet. Abschließend werden Ansätze zur Entwicklung von Spracherkennungstechnologien inklusive Datenbanken beschrieben, die insbesondere bei der Verwendung von medizinischem Vokabular eine unterstützende Rolle spielen.

Das vierte Kapitel legt die Methodik zur Beantwortung der Forschungsfragen dar. Weiters wird die Literaturrecherche laut dem PRISMA-Standard analysiert. Anschließend wird der Verlauf der Erstellung des Fragebogens, der Verlauf der Beobachtungsanalyse und des Interviews beschrieben.

Im fünften Kapitel wird zunächst ein Überblick über das Produkt "Dragon Medical SpeechKit" der Firma Nuance gegeben und darüber hinaus die Konkurrenz am Markt vorgestellt. Weiters werden Erfahrungsberichte zusammengefasst, um einen Eindruck über die Zufriedenheit der Benutzer zu erhalten. Danach wird speziell auf den Einsatz bei den Barmherzigen Brüdern Graz eingegangen und die Einführung in den klinischen Workflow mithilfe der im Interview erlangten Informationen beschrieben. Abschließend folgt die Auswertung des Fragebogens und der Beobachtungsanalyse.

Im sechsten Kapitel werden die Ergebnisse der Arbeit zusammengefasst und diskutiert. Abschließend wird auf die Zukunft der Spracherkennung in der medizinische Dokumentation und auf zukünftige Trends des Gesundheitswesens eingegangen.

Die Schlussfolgerung fasst die wichtigsten Ergebnisse zusammen und gibt einen Ausblick auf die nächsten Schritte der Spracherkennung im Gesundheitswesen.

Kapitel 2

Spracherkennung

Spracherkennung steht im Zentrum der Interaktion zwischen Menschen und Technologie, indem sie gesprochene Sprache in Text umwandelt und somit laut Sosnowsky [2023] eine natürlichere Kommunikation mit Maschinen ermöglicht. Diese Technologie, die seit über siebenzig Jahren entwickelt wird, hat bedeutende Fortschritte erzielt und ist heute ein integraler Bestandteil vieler Anwendungen, die von der Dokumentation bis hin zur künstlichen Intelligenz reichen. Trotz der Vielfalt und Komplexität menschlicher Sprache, einschließlich Akzente, Dialekte und Sprechgeschwindigkeit, arbeiten Forscher:innen laut Blackley et al. [2019] kontinuierlich daran, die Genauigkeit und Echtzeitübersetzung von Spracherkennungssystemen zu verbessern.

2.1 Definition und Grundprinzipien

Spracherkennung ist laut Carstensen et al. [2010] ein Verfahren der angewandten Informatik und Computerlinguistik. Es befasst sich mit der Untersuchung und Erforschung von dem Verfahren, gesprochene Sprache für die Datenerfassung zugänglich zu machen. Die menschliche Sprache ist vielfältig, mit allen möglichen Akzenten und Dialekten, deswegen müssen Spracherkennungstechnologien auch flexibel sein und Anpassungsmöglichkeiten, wie verschiedene Modelle, mit sich bringen. Die beiden wichtigsten Aspekte eines Spracherkennungssystems sind laut Malik et al. [2021] die Genauigkeit des Ergebnisses und die Geschwindigkeit des Umwandlungsprozesses. Während die Genauigkeit und Geschwindigkeit der Spracherkennungstechnologie von zentraler Bedeutung sind, wirft der Versuch, die subtilen Nuancen und die Dynamik der menschlichen Stimme zu erfassen, einzigartige technische Herausforderungen auf. Die Komplexität dieses Prozesses wird deutlich, wenn man den menschlichen Stimmapparat und den Prozess der Lauterzeugung betrachtet. Der menschliche Stimmapparat besteht laut Lee et al. [2021] aus drei Teilen: der Lunge, dem Kehlkopf und dem Sprechapparat. Der Vorgang beginnt in der Lunge. Nachdem die Luft

die Luftröhre passiert hat, erzeugt die Lunge einen Luftstrom. Der Luftdruck steigt in den Stimmbändern an, im Kehlkopf öffnet sich die Glottis, so nennt man laut Lenarz und Boeninghaus [2012] in der Anatomie den stimmbildenden Apparat, der aus Stimmlippen und Stellknorpel, sowie der dazwischen liegenden Stimmritze besteht. Die Stimmlippen beginnen in der Grundfrequenz (125-200 Hz) zu schwingen und erzeugen eine Reihe von Impulsen. Diese erregte Luft strömt durch die Nasenwege, den Mund und die Lippen und erzeugt einen wahrgenommenen Laut, die Sprache.

2.1.1 Arten von Spracherkennungstechnologien

Das Ziel der Spracherkennung (SE) besteht darin, gesprochene Sprache unabhängig vom Sprecher zu interpretieren. Genauso wie menschliche Stimmen sich voneinander unterscheiden, zeigen auch Systeme zur Spracherkennung diverse Viabilitäten. Spracherkennung kann laut Hopkins [2024] grob von zwei Richtungen betrachtet werden: die Front-End- und die Back-End-Spracherkennung. Die Front-End-Spracherkennung erfolgt in Echtzeit. Dabei sprechen die Benutzer:innen in ein Mikrofon, und das Gesprochene wird in Text umgewandelt. Dieser Ansatz wird oft für Anwendungen wie Echtzeit-Transkription und Sprachbefehle verwendet. Bei Fehlern wie Rechtschreibfehlern oder fehlenden Wörtern ist jedoch menschliches Eingreifen erforderlich, um die Genauigkeit sicherzustellen. Die Back-End-Spracherkennung hingegen erfolgt asynchron. Hierbei werden Sprachaufnahmen, die mit einem Mikrofon oder einem Diktiergerät gemacht wurden, zunächst als digitale Audiodateien gespeichert. Diese Audiodateien werden dann später verarbeitet, um das Gesprochene in Text umzuwandeln. Diese Methode wird häufig in professionellen Umgebungen wie medizinischen oder juristischen Transkriptionen verwendet, wo menschliche Mitarbeiter die Audiodateien überprüfen und manuell in Text umwandeln. Über die grundlegenden Unterscheidungen hinaus spielt auch die Verarbeitung der Sprecherdaten eine entscheidende Rolle, es gibt drei verschiedene Arten:

Sprecher:innenunabhängig

Diese Systeme sind laut Rippel [2021] darauf ausgelegt, Sprache von verschiedenen Sprechern zu erkennen und funktionieren daher unabhängig. Sie sind nicht speziell auf einen einzelnen Benutzer zugeschnitten und erfordern in der Regel einen komplexeren Entwicklungsprozess. Obwohl diese Systeme möglicherweise eine geringere Genauigkeit im Vergleich zur sprecher:innenabhängigen Methoden aufweisen, bieten sie eine größere Flexibilität und eignen sich besser für den alltäglichen Gebrauch.

Sprecher:innenabhängig

Diese Systeme sind so entwickelt, dass sie durch ein oder mehrere Personen, deren Stimmprofil das System vorher erlernt hat, benutzt werden können. Laut Schukat-Talamazzini [2013] sind sie in der Regel leichter anzupassen und können oft eine bessere Leistung erzielen. Allerdings ist ihre Fähigkeit, mit Daten oder Stimmen außerhalb des spezifisch trainierten Benutzerkreises präzise zu arbeiten, eingeschränkt.

Hybridsysteme

Diese Systeme liegen laut Malik et al. [2021] in etwa zwischen der Sprecher:innenunabhängigen und Sprecher:innenabhängigen. Diese Systeme werden so trainiert, dass sie neue Sprachmuster erlernen können, wenn ein:e neuer Sprecher:in auftritt. Sie bieten somit die Vorteile beider Modelle, die Genauigkeit des trainierten Sprachprofils der Benutzer:innen und die Anpassungsfähigkeit an neue Benutzer:innen.

2.1.2 Vokabular

Zusätzlich gibt es laut Jongman et al. [2021] Unterschiede in der Größe des Vokabulars einer SE, was sich auf die Komplexität, die Verarbeitungszeit und die Genauigkeit auswirken kann. Je größer der Wortschatz, desto komplexer das System [Malik et al., 2021]. Einige Systeme, wie zum Beispiel eine Spracherkennung für gesprochene Zahlen von null bis neun, benötigen nur zehn Vokabeln, während für eine Spracherkennung von menschlicher Sprache tausende nicht ausreichend sein können.

Weiters wird zwischen isoliertem und verbundenen Vokabular unterschieden. Ein direktes System, das auf isolierten Wörtern basiert, erfordert laut Tscheschner [2007], dass die Benutzer:innen zwischen jedem Wort eine ausdrucksvolle Pause machen. Dies bedeutet nicht notwendigerweise, dass das System nur ein Wort auf einmal verarbeiten kann, sondern dass das System mehrere Vokabeln auf einmal aufnehmen kann, diese aber einzeln verarbeitet. Bei zusammenhängenden Vokabularen handelt es sich um ein kontinuierliches System, das mit zusammenhängenden Wortketten arbeitet und keine Pausen von Benutzer:innen verlangt. Solche Systeme können die aufgenommenen Wortketten anschließend in einem Schritt verarbeiten.

2.1.3 Sprachmodelle

Darüber hinaus kann zwischen verschiedenen Modellen unterschieden werden, die am Ende der Konvertierung verwendet werden. Es gibt laut Joshi et al. [2017] drei Modelle: das

akustische Modell, das Aussprachmodell und das Sprachmodell. Das akustische Modell bildet die Verbindung zwischen den linguistischen Phonemen und dem Audio-Signal ab und ermöglicht so, zwischen Stimme und Lauten zu unterscheiden. Das Aussprachmodell definieren die Kombinationsregeln der Phoneme zur Bildung von Wörtern. Das Sprachmodell hingegen analysiert und vergleicht die Lautsequenzen, um ähnlich klingende Wörter voneinander zu trennen.

2.1.4 Umwandlung

Die zuvor erläuterten Begriffe sind bei der Umwandlung von Sprache in Text zu berücksichtigen. Laut Joshi et al. [2017] lässt sich der Umwandlungsprozess in sieben Module unterteilen.

1. *Empfangen eines Audiosignals*: Die Tonaufnahme ist der erste Schritt der Spracherkennung. Sie kann mit üblichen Mikrofonen durchgeführt werden. Anschließend werden die Schallwellen in ein digitales Format umgewandelt, damit sie von einem Computer verarbeitet werden können. Dabei wird das Audiosignal in seine akustischen Grundmerkmale zerlegt. Dazu gehören Tonhöhe, Lautstärke und Dauer.
2. *Audio-Vorverarbeitung*: Die empfangenen Audiodaten werden zunächst einer Vorverarbeitung unterzogen, um störende Hintergrundgeräusche zu minimieren und die Klarheit der Benutzersprache zu verbessern. Dadurch wird die Qualität der Spracherkennung deutlich erhöht. In einem weiteren Schritt werden die akustischen Merkmale präzise in phonetische Einheiten umgewandelt. Die gründliche phonetische Analyse spielt eine entscheidende Rolle, denn sie ermöglicht die korrekte Identifikation von Wörtern und die Erfassung ihrer Bedeutung und schafft damit eine verlässliche Basis für das Verständnis und die Verarbeitung der Sprachdaten.
3. *Zerlegung in Fragmente*: Die Audiodatei wird in kleine Fragmente von zehn bis 25 Millisekunden zerlegt. Dies ist notwendig, um die Analyse der Audiodaten zu optimieren. Jedes Fragment wird einzeln analysiert.
4. *Extraktion von Sprachmerkmalen*: In dieser Phase werden Sprachparameter wie Frequenz, Dauer und Amplitude berechnet. Sie werden verwendet, um die Phoneme (Laute/Wörter) zu identifizieren, aus denen die gesprochenen Wörter bestehen. Hier werden bei traditionellen Systemen Modelle und Algorithmen verwendet, die Muster in der Sprache erkennen und entsprechend umwandeln. Bei Systemen, die künstliche Intelligenz verwenden, werden laut Thakur et al. [2023] Ansätze aus dem maschinellen Lernen und deren Untergruppe Deep Learning verwendet, um direkt die Muster

aus der Audiodatei zu lernen, ohne auf vordefinierte Muster oder Algorithmen angewiesen zu sein.

5. *Vergleich mit Erkennungsmodellen*: Die Phoneme werden dann mit Modellen verglichen, die mit Sprachproben trainiert wurden. Die in 2.1.3 beschriebenen Sprachmodelle verwenden verschiedene Ansätze, die jeweils aus großen Sprachdatenbanken lernen, um die Erkennungsleistung zu verbessern. Auch die Anpassung an einzelne Sprecher:innen oder branchenspezifisches Vokabular kann in diesem Schritt erfolgen. Bei Systemen mit künstlicher Intelligenz kann dieser Schritt übersprungen werden.
6. *Komposition von Wörtern und Text*: Das Sprachmodell kombiniert Phoneme zu Wörtern und Sätzen und schließlich zu vollständigem Text. Bei Systemen, die künstliche Intelligenz verwenden, werden hier die erkannten Muster zu Wörtern und Sätzen zusammengestellt.
7. *Einfügen und Überarbeitung*: Nachdem der von Audio umgewandelte Text in das Dokument eingefügt wurde, hat der/die Benutzer:in die Möglichkeit, Fehler zu korrigieren, die während der Umwandlung aufgetreten sind. Dieser Prozess ermöglicht es den Benutzer:innen, eine genaue und fehlerfreie Version des Textes zu erstellen, die ihren Anforderungen und Standards entsprechen.

2.1.5 Qualitätskriterien

In der Welt der Spracherkennungstechnologien können zwei Qualitätskriterien zur Bewertung von Spracherkennungssystemen herangezogen werden: die Genauigkeit und die Geschwindigkeit des SE-Systems. Aus diesen Werten kann dann auf die Effizienz geschlossen werden.

Die Geschwindigkeit kann laut Malik et al. [2021] durch die Berechnung der Verarbeitungsgeschwindigkeit bestimmt werden. Diese kann mit folgender Formel berechnet werden:

$$\frac{P}{I} \tag{2.1}$$

Dabei ist P die Zeit, die das System zur Verarbeitung der Eingabe benötigt, und I die Dauer der Eingabe.

Die Wortfehlerrate dient laut Jurafsky und Martin [2024] als Maßstab für die Präzision des Systems. Sie hilft, die Effizienz der Spracherkennungstechnologie zu bewerten, indem sie die Genauigkeit auf Wortebene, im Gegensatz zur Geräuschebene beurteilt. Die Berechnung der Wortfehlerrate ergibt sich aus:

$$\frac{S + D + Q}{N} \quad (2.2)$$

Dabei steht S für die Gesamtzahl der Änderungen, die im Vergleich zum Originaltext im Ausgangstext vorgenommen wurden. D steht für die Anzahl der Streichungen, während Q die Anzahl der Einfügungen im Text angibt. N ist die Gesamtzahl der Wörter im Text.

2.2 Geschichte der Spracherkennung

Das Interesse an der Spracherkennung reicht bis ins frühe 20. Jahrhundert. In den 1930er Jahren schlug Homer Dudley von den Bell Laboratories laut Raphael et al. [2021] ein Systemmodell für die Analyse und Synthese von Sprache vor. Im Jahr 1952 entwickelte dieselbe Organisation laut Pieraccini [2012] das System "Audrey", das in der Lage war, eine einzelne (männliche) Stimme zu erkennen, welche die Zahlen von eins bis zehn laut aussprach. Zehn Jahre später brachte die International Business Machines Corporation (IBM) laut Bajpai und Sharda [2022] "Shoebox" auf den Markt, das 16 englische Wörter verstand und darauf reagierte konnte. Weltweit wurde weitere Hardware entwickelt, die Töne und Sprache erkennen konnte. Ende der 1960er Jahre waren Spracherkennungstechnologien laut Xiong [2022] in der Lage, Wörter mit vier Vokalen und neun Konsonanten zu verarbeiten. In den 1970er Jahren wurden die Fortschritte laut Pallett [1989] vom US-Verteidigungsministerium und der DARPA (Defense Advanced Research Projects Agency) vorangetrieben. Das aus diesem Programm hervorgegangene Sprachsystem "Harpy" der Carnegie Mellon University konnte laut Lowerre [1976] mehr als tausend Wörter verstehen, was in etwa dem Wortschatz eines dreijährigen Kindes entspricht. In den 1980er Jahren machte SE einen bedeutenden Fortschritt in der statischen Sprachmodellierung. Einer der Durchbrüche war laut Xuedong et al. [2014] das "Hidden Markov Modell". Anstatt nur Wörter zu verwenden und nach Klangmustern zu suchen, schätzte das "Hidden Markov Model (HMM)" die Wahrscheinlichkeit, dass bestimmte unbekannte Laute bestimmte Wörter sind. Darüber hinaus brachte IBM das Spracherkennungssystem "Tangora" Bahl et al. [1988] auf den Markt, das eine Vokallänge von zwanzigtausend Wörtern verarbeiten konnte. 1982 gründeten James und Janet M. Baker die Firma Dragon Systems [Xuedong et al., 2014], die zu einem der wenigen Konkurrenten von IBM wurde.

In den 1990er Jahren wurde die Spracherkennung vor allem durch Laptops vorangetrieben. Schnellere Prozessoren ermöglichten den weit verbreiteten Einsatz von Software wie "Dragon Dictate" [Mandel, 1992]. BellSouth führte "VAL" [Sokolov et al., 2020] ein, ein

interaktives Spracherkennungssystem, das über Telefonnummernwahl funktioniert. Dieses System bildete die Grundlage für Telefonbaumsysteme, die auch heute teilweise in den USA noch verwendet werden.

In den frühen 2000er Jahren unterstützte DARPA die Spracherkennungsforschung durch Programme wie EARS (Effective Affordable Reusable Speech-to-text) [Cieri et al., 2004] und GALE (Global Autonomous Language Exploitation) [Olive et al., 2011]. Während dieser Zeit dominierten traditionelle Ansätze die Forschung, aber gegen Ende des Jahrzehnts führte die Einführung von Deep Learning zu signifikanten Fortschritten. Trotz der Tatsache, dass in den meisten Jahren dieses Jahrzehnts nur geringe Fortschritte erzielt wurden, markierte die Einführung von "Google Voice Search" laut Jyothi et al. [2012] einen Wendepunkt. Dieses Tool ermöglichte es Google, Milliarden von Daten aus Suchanfragen zu sammeln, was die Vorhersage verbesserte, nach welchen Informationen Nutzer suchen könnten. Insgesamt sammelte Google 230 Milliarden Wörter durch das Voice Search-Programm. Zusätzlich startete Google im Jahr 2007 einen telefonbasierten Verzeichnisdienst.

Die 2010er Jahre waren geprägt von einem Durchbruch mit der Einführung von Apples "Siri" [Hoy, 2018], gefolgt von einer Flut neuer Spracherkennungsanwendungen. Zudem traten Smart-Home-Systeme wie Amazons "Alexa" und "Google Home" [Li et al., 2017] in den Markt ein. Diese Entwicklungen löste laut Sonix [2024] einen intensiven Wettbewerb zwischen großen Technologieunternehmen wie Microsoft, IBM und Google aus.

2.3 Branchen

Spracherkennung wird laut Kumar [2024] in verschiedenen Bereichen eingesetzt, von virtuellen Assistenten über Transkriptionssysteme bis hin zur Sprachsteuerung im Auto. Der folgende Abschnitt zeigt, in welchen Bereichen Spracherkennung eingesetzt wird:

- *Kfz-Systeme*: Fahrzeugsysteme ermöglichen die Aktivierung der Spracherkennung, zum Beispiel durch Drücken einer Taste am Lenkrad, wodurch ein "Hörfenster" für Spracheingaben geöffnet wird und die Benutzer:innen anschließend ihren Wunsch äußern können.
- *Gesundheitswesen*: Im Gesundheitswesen erleichtert Spracherkennung die medizinische Dokumentation, entweder direkt durch Diktat in ein Spracherkennungssystem (Front-End) oder durch nachträgliche Bearbeitung erkannter Sprache (Back-End).

- *Militär*: Im militärischen Bereich verbessert die Spracherkennung die Effizienz in Kampfflugzeugen, von der Steuerung über die Kommunikation bis hin zur Waffeneinsatzplanung.
- *Kommunikation*: In der Telekommunikation ermöglicht sie eine effizientere Kundenbetreuung und kürzere Wartezeiten. Dies führt zu einer höheren Kundenzufriedenheit.
- *Bildung*: Im Bildungsbereich unterstützt Spracherkennung, das Erlernen neuer Sprachen und hilft Menschen mit Behinderungen, Computer über Sprachbefehle zu bedienen.

Darüber hinaus können Spracherkennungstechnologien in Bereichen wie intelligenten Heimsystemen, Bankwesen und Übersetzungsdiensten eingesetzt werden.

2.4 Ansätze

Traditionelle SE-Systeme verwenden laut Thakur et al. [2023] regelbasierte Modelle und statistische Modelle, um gesprochene Sprache in Text umzuwandeln. Diese Systeme basieren auf vordefinierte Algorithmen, die darauf abzielen, Muster in der Sprache zu erkennen und diese entsprechend umzuwandeln. SE-Systeme, die mit künstlicher Intelligenz (KI) arbeiten, nutzen Ansätze aus der Untergruppe von KI, dem maschinellen Lernen und dessen Untergruppe Deep Learning.

2.4.1 Maschinelles Lernen

Insbesondere seit den 2000er Jahren hat der Bereich der künstlichen Intelligenz (KI), einschließlich des maschinellen Lernens (ML), an Aufmerksamkeit gewonnen. Maschinelles Lernen ist, laut Rebala et al. [2019], ein Kerngebiet der Informatik, und beschäftigt sich mit der Entwicklung von Algorithmen und Techniken zur Automatisierung komplexer Probleme, die mit herkömmlichen Programmiermethoden nur schwer zu lösen sind. ML-Algorithmen zeichnen sich dadurch aus, dass sie aus datengestützten Erfahrungen lernen, anstatt auf vordefinierte, Anweisungen angewiesen zu sein. Innerhalb des maschinellen Lernens werden verschiedene Modelle verwendet, darunter das Hidden Markov Model (HMM), das Gaussian Mixture Model (GMM), Neuronale Netze (NN), Support Vector Machines (SVM), K-Nearest Neighbors (KNN) und Entscheidungsbäume(DT).

- *Hidden Markov Models (HMMs)*: sind ein leistungsfähiges Werkzeug für verschiedene Spracherkennungsaufgaben, insbesondere für die automatische Spracherkennung,

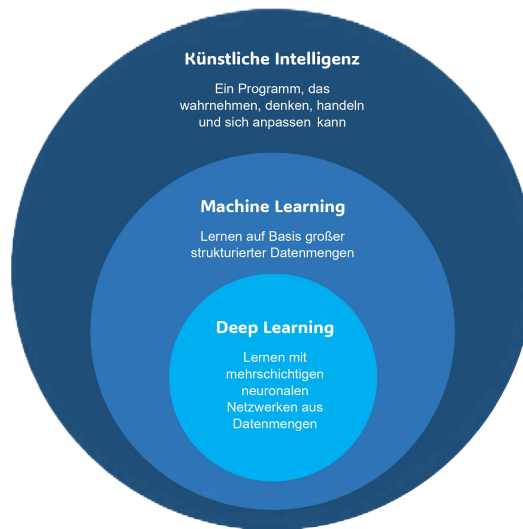


Abbildung 2.1: Visualisierung der Hierarchie zwischen künstlicher Intelligenz, maschinellen Lernens, und Deep Learning, dargestellt durch sich überlappende Kreise, die zeigen, wie Deep Learning ein spezifischer Teil von maschinellen Lernens ist, das selbst eine Untergruppe der künstlichen Intelligenz darstellt [datasolut, 2024]

dies beschreibt SE, die künstliche Intelligenz in der Umwandlung verwendet. HMMs werden laut Malik et al. [2021] eingesetzt, um zu analysieren, wie wahrscheinlich bestimmte Sprachlaute auftreten. Sie tun dies, indem sie eine Abfolge von nicht direkt sichtbaren Zuständen verwenden. Diese Zustände sind mit dem, was wir hören können, über Beobachtungen verbunden. Einfach ausgedrückt helfen HMMs dabei zu verstehen, in welcher Reihenfolge Sprachlaute typischerweise erscheinen, indem sie die verborgenen Muster hinter diesen Lauten aufdecken.

- *Gaussian mixture models (GMMs)*: Das GMM-Modell wird laut Deepa und Khilar [2022] verwendet, um die Wahrscheinlichkeitsverteilung von kontinuierlichen Merkmalen zu simulieren und eignet sich gut zur Modellierung von Sprachmerkmalen. Es wird benutzt, um Patienten mit verschiedenen Stimmstörungen zu klassifizieren. Es zeigte sich, dass GMM effektiv analysieren kann, indem es Merkmalsextraktionsmethoden verwendet, die wichtige Aspekte wie Konsonanten- und Vokalpositionen berücksichtigen.
- *Neuronale Netze (NN)*: sind inspiriert von der Struktur menschlicher Nervensysteme und nutzen künstliche Neuronen zur Datenverarbeitung. Sie sind laut Deepa und Khilar [2022] besonders wirksam in der Erkennung und Klassifizierung von Mustern, was sie für die Anwendungen, in der Spracherkennung und medizinischen Diagno-

stik wertvoll macht. Durch Techniken wie das Multilayer Perceptron (MLP) und Vektorquantisierung, laut Alanazi et al. [2017] ein Verfahren zu Datenkompression und Mustererkennung, sind sie effizient in der Analyse spezifischer Merkmale wie Sprachfehler.

- *Support vector machines (SVMs)*: sind laut Thakur et al. [2023] eine leistungsstarke Klasse von überwachten Lernalgorithmen und eignen sich hervorragend für eine Vielzahl von Sprachklassifikationsaufgaben, wie Sprecheridentifikation und Phonemerkennung. SVMs stellen Daten als Punkte in einem Raum dar und suchen die beste Linie oder Fläche, die verschiedene Gruppen trennt. Diese Technik ist sehr gut darin, unterschiedliche Arten von Daten zu unterscheiden und vorherzusagen, zu welcher Gruppe neue Daten gehören.
- *K-nearest neighbors*: Dieser Algorithmus stellt laut Deepa und Khilar [2022] einen einfachen, aber effektiven Ansatz für Klassifikationsaufgaben dar, der sich in einer Vielzahl von sprachbezogenen Anwendungen bewährt hat, einschließlich Sprecheridentifikation und Spracherkennung. Das Grundprinzip ist, für bestimmte Daten die nächsten Beispiele in bereits bekannten Daten zu finden und diese neuen Daten der Gruppe zuzuordnen, die unter diesen nahen Beispielen am häufigsten vorkommt.
- *Entscheidungsbäume*: Sie strukturieren den Merkmalsraum durch rekursive Unterteilung und erstellen Entscheidungsregeln für die Zuordnung von Datenpunkten zu Klassen. Diese Methode ermöglicht es Entscheidungsbäumen (DT), laut Alanazi et al. [2017] komplexe Entscheidungsgrenzen effizient abzubilden. DTs klassifizieren neue Daten, indem sie diese von der Wurzel des Baumes bis zu einem Endknoten leiten, basierend auf den erstellten Regeln.

2.4.2 Deep Learning

Deep Learning ist laut Chassagnon et al. [2020] ein Bereich des maschinellen Lernens, der auf der Verwendung tiefer neuronaler Netzwerke basiert. Diese Netzwerke sind im Wesentlichen mehrschichtige Strukturen von Neuronen, die die Fähigkeit besitzen, aus großen Datenmengen direkt zu lernen. Im Unterschied zu einfacheren neuronalen Netzwerken, wie dem Multilayer Perceptron, die nur über wenige Schichten verfügen, können tiefe neuronale Netzwerke durch ihre vielen Schichten komplexe Muster in Daten erkennen und interpretieren. Zu Modellen von Deep Learning zählen: Tiefe neuronale Netze (DNN), Convolutional neural network (CNN), Convolutional deep belief network (CDBN), Recurrent Neural Networks (RNN).

- *Tiefe neuronale Netze (DNN)*: Sie erweitern herkömmliche KNN durch zusätzliche Schichten, um laut Deepa und Khilar [2022] komplexere Aufgaben zu lösen und eine genauere Verarbeitung von Eingaben zu Ausgaben zu ermöglichen. Durch diese Erweiterung können DNN anspruchsvollere Funktionen abbilden. Insbesondere bei der Erkennung von Stimmstörungen zeigen DNNs laut Deepa und Khilar [2022] ihr Potenzial durch eine hohe Genauigkeit bei der Analyse verschiedener Stimmpathologien.
- *Convolutional neural network (CNN)*: war laut Xiong [2022] ein Durchbruch, da es das Erste Deep-Learning-Modell mit einer mehrschichtigen neuronalen Netzwerkstruktur war. Ursprünglich für die Bildverarbeitung konzipiert, entdeckten Forscher bald seine Leistungsfähigkeit in der Spracherkennung. Durch die Integration einer Pooling-Schicht können mehrere ähnliche Sprachsignale zu einem einzigen Signal zusammengefasst werden, was den Rechenaufwand bei der Verarbeitung natürlicher Sprache erheblich reduziert. Darüber hinaus erlaubt die Anwendung der Faltungsschicht dem CNN Kontextinformationen zu extrahieren. Sie unterscheiden sich von CDBNs durch ihre Fähigkeit, ähnliche Daten zu aggregieren und Kontextinformationen zu extrahieren, was sie in spezifischen Anwendungsbereichen wie der Spracherkennung leistungsfähig macht.
- *Convolutional deep belief network (CDBN)*: Als Feedforward-Netzwerke basieren sie auf drei Schlüsselkonzepten, Faltung, Downsampling und effiziente Datenverarbeitung, um Robustheit gegenüber Verschiebungen und Deformationen zu gewährleisten. CDBNs nutzen laut Xiong [2022] Faltung, um wichtige Merkmale in den Sprachsignalen, wie bestimmte Laute oder Intonationen, zu identifizieren. Das Downsampling hilft dann dabei, diese Informationen auf eine handhabbare Größe zu reduzieren, ohne wichtige Details zu verlieren. So können CDBNs effektiv lernen, welche Klangmuster bestimmten Worten oder Phrasen entsprechen. Im Gegensatz zu CNNs, die auf komplexe Strukturen mit zusätzlichen Schichten setzen, fokussieren sich CDBNs auf eine robuste und effiziente Datenverarbeitung mit minimalen Vorverarbeitungsanforderungen.
- *Recurrent Neural Networks (RNN)*: künstlichen Neuronen sind meistens so aufgebaut, dass sie entweder Verbindungen zu sich selbst oder zu vorherigen Ebenen haben, was ihnen eine Art Kurzzeitgedächtnis verleiht. Diese Struktur des RNN ermöglicht es dem Netzwerk laut Song [2020], sich an frühere Daten zu erinnern und diese in seine aktuellen Entscheidungen mit einzubeziehen. Diese Netzwerke bekannt als "Recurrent Neural Networks"-sind speziell dafür entwickelt, Daten zu

verarbeiten, die in Sequenzen angeordnet sind, wie zum Beispiel Texte oder Zeitreihen. Sie lösen komplexe Probleme, indem sie Sequenzen als Abfolgen von Ereignissen betrachten und nutzen dafür zwei Hauptmerkmale: Persistenz und Gedächtnis. Persistenz bedeutet, dass das Ergebnis eines aktuellen Moments auch von früheren Daten abhängt. Und Gedächtnis, das Netzwerk, Informationen über mehrere Zeitpunkte hinweg zu speichern und zu nutzen, um Muster oder Abhängigkeiten in den Daten zu erkennen und darauf basierend Entscheidungen zu treffen.

Kapitel 3

Spracherkennung im Gesundheitswesen

Spracherkennung spielt im Gesundheitswesen eine wichtige Rolle, da Ärzt:innen etwa sechzig Prozent ihres Arbeitstages mit Dokumentation verbringen und somit nur vierzig Prozent der Zeit für den Patienten bleibt. Die Technologie hat laut Kumar [2024] das Potenzial, die Effizienz und Genauigkeit der Befundung erheblich zu verbessern, was nicht nur die Arbeitsbelastung des medizinischen Personals verringern, sondern auch die Qualität der Versorgung verbessern kann.

3.1 Medizinische Befunderhebung

Die medizinische Befunderhebung bildet das Herzstück des diagnostischen Verfahrens im Gesundheitssektor und ist essenziell für die Gewährleistung einer präzisen und effektiven Patientenversorgung. Dieser Prozess beinhaltet laut Kumar [2024] die umfassende Sammlung und detaillierte Analyse von Patientendaten, die entscheidend sind, um den gegenwärtigen Gesundheitszustand eines Individuums zu beurteilen, sowie mögliche Krankheitsbilder zu identifizieren.

Der Prozess der Befunderhebung beginnt laut Gressner und Gressner [2019] in der Regel mit der Anamnese. In diesem initialen Schritt führen Ärztinnen und Ärzte ein tiefgehendes Gespräch mit dem Patienten, um dessen medizinische Vorgeschichte, gegenwärtige Symptome und bestehende Beschwerden zu ermitteln. Diese Informationen sind grundlegend, da sie den medizinischen Fachkräften erste Hinweise auf mögliche gesundheitliche Probleme geben.

Nach der Anamnese folgen körperliche Untersuchungen, bei denen der Ärzt:innen den Körper des Patienten auf Anzeichen von Krankheiten überprüft, die mit den berichteten Symptomen in Verbindung stehen könnten. Abhängig von den Ergebnissen dieser ersten Untersuchungen können spezifische diagnostische Tests angeordnet werden. Diese Tests umfassen laut Gressner und Gressner [2019] ein breites Spektrum an Verfahren, von

Blutuntersuchungen, über bildgebende Verfahren, bis hin zu Biopsien und dienen dazu, die vorläufigen Diagnosen zu bestätigen oder zu widerlegen. Innerhalb der medizinischen Befunderhebung können die Ergebnisse vielfältig sein und zum Beispiel den Blutzuckerspiegel oder bestimmte Verhaltensweisen des Patienten umfassen. Es existieren laut Haas [2005] vier primäre Arten von medizinischen Befunden:

1. *Gesamtbefund*: Die Zusammenfassung aller gesammelten Einzelbefunde.
2. *Pathologischer Befund*: Bezieht sich auf die Identifizierung von Krankheitsbildern oder pathologischen Veränderungen.
3. *Histologischer Befund*: Involviert die Analyse von Gewebeproben unter dem Mikroskop, um Zellstrukturen und -anomalien zu identifizieren.
4. *Psychopathologischer Befund*: Fasst die Ergebnisse von psychologischen oder psychiatrischen Bewertungen zusammen und liefert Einblicke in die psychische Verfassung des Patienten.

Die korrekte Erstellung und präzise Interpretation von Befunden zählen zu den grundlegenden Prozessen im ärztlichen Alltag. Diese Kompetenzen sind laut Dettmeyer [2006] besonders wichtig für die sachgerechte Dokumentation von medizinischen Leistungen zu Abrechnungszwecken, sowie in rechtlichen Zusammenhängen, beispielsweise bei gerichtlichen Verfahren. Zusammengefasst ermöglicht eine sorgfältige und detaillierte Befunderhebung dem medizinischen Fachpersonal, fundierte Diagnosen zu stellen und darauf basierend effektive Behandlungsstrategien zu entwickeln.

3.2 Meilensteine der Vergangenheit

Schon vor dem digitalen Wandel in der Gesundheitsbranche wurden laut Ronaldo et al. [2004] mehrere Innovationen eingeführt, um die Dokumentation der Patientendaten zu verbessern. Die Schreibmaschine war laut Darnell [1997] ein erster Versuch, diesem Ziel näherzukommen, gefolgt von Großrechnern der 1950er Jahre.

Die Wurzeln der Spracherkennung in der Medizin reichen laut Kudryavtsev et al. [2023] bis in die 1970er und 1980er Jahre zurück, als die ersten Studien über Technologien zur Erzeugung und Nutzung von geschriebenem oder gesprochenem Text veröffentlicht wurden, die unter Begriffen wie "Computational Linguistics", "Natural Language Processing" und "Text Mining" bekannt wurden. 1981 wurde ein bedeutender Fortschritt erzielt, als das Beth Israel Hospital das CLIP-System (Coded Language Information Processing) [Kim und Leeming, 1990] einführte, das in der Lage war, Patientenakten in einer kodierten medizinischen Ontologie, ein Framework um medizinisches Wissen systematisch zu organisieren, zu klassifizieren, obwohl es durch einen kleinen Arbeitsspeicher und ein begrenztes Vokabular von etwa zweihundert medizinischen Vokabeln eingeschränkt war. In den 1990er Jahren wurden laut Johnson et al. [2014] Audio-Transkriptionszentren eingerichtet, um das wachsende Arbeitsvolumen der Transkriptionist:innen zu bewältigen. Medizinisches Personal sprach Befunde in Tonbandgeräte, die dann durch Transkriptionist:innen in Text umgewandelt wurden. Zusätzlich begann das amerikanische Gesundheitswesen laut Fishman [1997] 1994 Spracherkennungstechnologien für medizinische Berichte zu implementieren. Doch die Herausforderung war laut Hoffman [1997], dass die ersten Systeme verlangten, dass die Ärzt:innen lernten, mit dem Computer zu sprechen, anstatt, dass sich der Computer an die Ärzt:innen anpasst. Die langsame Verarbeitungszeit der Computer dieser Zeit kam noch hinzu, sowie die Tatsache, dass nicht zwischen Hintergrundgeräusch und Sprache unterschieden werden konnte. Außerdem waren die Kosten hoch, das "Voice MedSystem" von Kurzweil kostete laut Fishman [1997] fast 25.000 Euro pro Arbeitsplatz. In den weiteren Jahren fand laut Ronaldo et al. [2004] eine stetige Verbesserung statt, der Wortschatz wuchs, die Anpassung an den medizinischen Fachjargon nahm zu und die Genauigkeitsraten stiegen an.

Zu Beginn der 2000er Jahre begann laut Gaikwad et al. [2010] die Einführung fortschrittlicherer Spracherkennungssysteme im Gesundheitswesen, die durch verbesserte Algorithmen und größere Datensätze gekennzeichnet waren. Ein Beispiel hierfür ist "Dragon Naturally Speaking Medical" [Nuance, 2010] von Nuance. Dieses System, speziell für medizinische Anwendungen entwickelt, ermöglichte es Ärzt:innen, Befunde und Patientenakten direkt in ihre Computer einzusprechen. Weiters wurde durch die Verbreitung von Cloud-Technologien in den 2010er Jahren Spracherkennungssysteme laut Muhammad [2015] zunehmend cloud-basiert, was die Zugänglichkeit und Skalierbarkeit dieser Technologien ohne umfangreiche Vorabinvestitionen in Hardware verbesserte. Ein Beispiel hierfür ist laut Everett et al. [2023] „Dragon Medical One“ von Nuance, das gegen Ende der 2010er Jahre populär wurde. Diese Plattform erlaubt es medizinischem Fachpersonal, medizinische Dokumentationen in Echtzeit zu diktieren. Dank Cloud-Technologie bietet das System hohe

Skalierbarkeit und kontinuierliche Updates, die die Genauigkeit und Benutzerfreundlichkeit stetig verbessern. Darüber hinaus hat die Integration von künstlicher Intelligenz (KI) laut Benkerzaz et al. [2019] in Spracherkennungstechnologien die Genauigkeit und Effizienz dieser Systeme weiter gesteigert. KI-Algorithmen lernen aus Interaktionen und passen sich im Laufe der Zeit an die spezifischen Bedürfnisse und den Fachjargon des medizinischen Personals an. Ein weiteres Beispiel hierfür ist laut Bajpai und Sharda [2022] "IBM" Watson, das KI nutzt, um große Mengen unstrukturierter Daten zu analysieren und medizinisches Personal durch die Bereitstellung relevanter medizinischer Informationen und klinischer Empfehlungen basierend auf den gesprochenen Anfragen zu unterstützen.

3.3 Anwendungsbereiche im Gesundheitswesen

Spracherkennungstechnologien haben sich als wertvolles Werkzeug in verschiedenen Bereichen des Gesundheitswesens etabliert, im folgenden Abschnitt werden diese beschrieben:

Medizinische Dokumentation

Die Nutzung von Spracherkennungstechnologien für die medizinische Dokumentation bietet eine effiziente Möglichkeit, medizinische Berichte und Patientenakten zu erstellen und zu aktualisieren. Dies umfasst:

- *Transkription von medizinischen Berichten:* Ärzt:innen und medizinisches Personal nutzen laut Goss et al. [2019] die Spracherkennung, um medizinische Berichte, Notizen und Befunde zu diktieren. Diese können auch automatisch in Patientenakten gespeichert werden.
- *Medizinische Dokumentationssysteme:* Dies ermöglicht es Ärzte:innen laut Goss et al. [2019] ohne Tastatur oder Maus auf medizinische Informationssysteme zuzugreifen und Patientenakten zu aktualisieren.
- *Pflegedokumentation:* Laut Everett et al. [2023] verwenden Pflegekräfte SE, um Pflegeprotokolle und -berichte effizient zu dokumentieren.
- *Medikationsmanagement:* Verordnungen und Medikationspläne können laut Pompili et al. [2020] per Spracheingabe dokumentiert oder abgerufen werden, was helfen kann, falsche Verordnungen zu minimieren und schneller zu entdecken.

Administrative Aufgaben und Kommunikation

- *Terminplanung*: Spracherkennung kann die Planung und Koordination von Terminen erleichtern, indem Termine per Sprachbefehl eingetragen werden können.
- *Dokumenten Archivierung*: Administrative Dokumente, wie Einwilligungserklärungen oder Versicherungsunterlagen können durch Diktat erstellt, kategorisiert und abgesichert werden.
- *Interne Kommunikation*: Sprachnachrichten können laut Le Prell und Clavier [2017] verwendet werden, um schnell Informationen zwischen Abteilungen oder Teammitgliedern auszutauschen, ohne schriftliche Berichte anfertigen zu müssen. Diese Berichte werden aber im Hintergrund geniert, damit die Dokumentation gegeben ist.

Patienteninteraktion

Spracherkennungstechnologien können die Interaktion mit Patient:innen verbessern und unterstützen das medizinische Personal bei:

- *Patienteninteraktion*: SE-Systeme können laut Walker et al. [2020] in interaktiven Anwendungen eingesetzt werden, um medizinische Informationen zu vermitteln, Fragen zu beantworten und Patienten über Behandlungspläne zu informieren.
- *Dolmetschen*: Echtzeitübersetzung gesprochener Informationen unterstützt laut Kuczera [2021] das medizinische Fachpersonal bei der Kommunikation mit Patienten, ohne dass ein externer Dolmetscher benötigt wird.
- *Stimmeurteilung*: Erfassung der Stimme des Patienten kann laut Sprötte [2022] zur Gewinnung wertvoller Hinweise über den Zustand des Patienten liefern.

Unterstützung im Operationsbereich

Die Integration von Spracherkennungstechnologien im Operationsbereich bietet Vorteile für die Durchführung und Dokumentation von Eingriffen sowie die Steuerung von assistierenden Robotersystemen;

- *Diktat von Operationen*: Chirurgen nutzen laut Aldosari et al. [2023] Spracherkennung, um während eines Eingriffs detaillierte Aufzeichnungen des Operationsverlaufs zu erstellen.
- *Roboterassistenzsysteme*: Ermöglichen laut Lobmeier [2020] die sprachgesteuerte Positionierung von Instrumente, zum Beispiel bei minimalinvasiven Operationen.

- *Notfallmedizin:* Spracherkennung kann laut Kim et al. [2021a] eine schnelle Erfassung von Informationen über den Zustand des Patienten in einer Notfallsituation liefern.

Spezialisierte medizinische Untersuchungen

In spezialisierten medizinische Fachrichtungen kann die Spracherkennungstechnologie praktisch eingesetzt werden, um Untersuchungen und Behandlungen effizienter zu gestalten.

- *Endoskopie:* In der Gastroenterologie kann SE laut Takayama et al. [2023] verwendet werden, um die Navigation von Endoskopiegeräten zu steuern. Dies ermöglicht den Ärzt:innen, sich nur auf die Untersuchung zu konzentrieren, während sie Befehle zur Steuerung des Geräts geben, ohne die sterile Umgebung zu gefährden.
- *Physiotherapeutische Übungen:* In der Physiotherapie können SE-Systeme laut [Pereira et al., 2019] eingesetzt werden, um Patienten durch Übungen zu führen. Die Systeme können Anweisungen geben, Feedback zur Ausführung sammeln und Anpassungen vorschlagen.
- *Bildgebung:* Radiologen können laut Kim et al. [2021b] SE-Befehle nutzen, um Bildgebungsgeräte zu steuern, Untersuchungssequenzen zu starten oder zu stoppen.
- *Pädiatrie:* Kinderärzte können interaktive sprachgesteuerte Spiele Ondáš et al. [2020] oder Tests nutzen, um Kinder spielerisch zu untersuchen. Dies kann helfen, Ängste abzubauen und eine kinderfreundliche Untersuchungsumgebung zu schaffen.
- *Geriatric:* SE kann genutzt werden, um ältere Patienten durch Programme zu führen, die laut [Do et al., 2020] auf Gedächtnisübungen und kognitive Tests zugeschnitten sind. Dies unterstützt die Bewertung und Überwachung des geistigen Zustands.
- *Anästhesie:* Anästhesisten können laut Jia und Yang [2022] während der Vorbereitung, Durchführung und Narkose die Spracherkennung nutzen, um Protokolle zu erstellen und wichtige Parameter zu dokumentieren. Dies ermöglicht eine lückenlose Überwachung des Patienten und verbessert die Patientensicherheit.

3.3.1 Datenbanken

Insbesondere bei der Spracherkennung im Gesundheitswesen sind spezialisierte Datensätze, die Sprach- und Audiosignale medizinischer Terminologie enthalten, entscheidend für die

Entwicklung von SE-Systemen. Diese Datensätze ermöglichen das Training von Machine Learning und KI-Systemen, um das domänenspezifische Vokabular des Gesundheitswesens zu erlernen. Solch ein zielgerichteter Ansatz verbessert die Genauigkeit der Spracherkennung in medizinischen Anwendungen erheblich.

- *Uncommon Voice*: Diese Datenbank beinhaltet laut Moore et al. [2020] Audioaufnahmen von Personen mit Sprachstörungen, die dazu dienen, Spracherkennungssysteme zu trainieren, um auch diese Nutzergruppe effektiv bei der Spracherkennung zu unterstützen.
- *Parkinson's disease speech dataset*: Diese Datenbank enthält laut Carrón et al. [2021] Sprachaufzeichnungen von Parkinson-Patienten. Sie werden verwendet, um Merkmale und Muster zu identifizieren, die bei der Früherkennung helfen können.
- *Medical Information Mart für Intensive Care III*: ist eine öffentliche Datenbank, die anonymisierte Gesundheitsdaten von Intensivpatienten enthält [MIMIC, 2021]. Sie enthält Herz- und Lungengeräusche sowie klinische Notizen und physiologische Kurven.
- *CommonVoice*: ist eine von Mozilla verwaltete öffentliche Datenbank, in der von Freiwilligen mehrsprachige Beispiele für 120 Sprachen gesammelt werden [Mozilla, 2024]. Dies hilft bei der Anpassung von SE-Systemen an verschiedene Sprachvarianten.

3.4 Vorteile und Herausforderungen

Vorteile

- *Beschleunigte Dokumentationsprozesse*: Ärzt:innen und medizinisches Personal können Befunde und Patientendaten mittels Spracherkennung eingeben, was den manuellen Eingabeaufwand eliminiert. Diese Praxis führt zu signifikanten Zeitersparnissen, die für den direkten Patientenkontakt genutzt werden können. Eine Studie von Zuchowski und Göller [2022] fand heraus, dass die durchschnittliche Dauer für die Erstellung medizinischer Dokumente mittels Spracherkennung bei 5,11 Minuten liegt, im Vergleich zu 8,9 Minuten für manuell eingegebene Dokumentation. Diese Beobachtung wird auch durch weitere Studien [Blackley et al., 2020], [Magrabi et al., 2018], [Kumar, 2024], [Goss et al., 2019] unterstützt.

- *Erhöhte Effizienz*: Durch die Nutzung von Spracherkennung können Informationen schneller erfasst und abgerufen werden, was den Zeitaufwand für die Verwaltung von Patientenakten reduziert. Das Paper "Physician use of speech recognition versus typing in clinical documentation: A controlled observational study" [Blackley et al., 2020] berichtet, dass Ärzt:innen die Anwendung von Spracherkennungssystemen als zeitsparend und effizient betrachten, da es ihnen ermöglicht schneller und mehr zu dokumentieren. Diese Einschätzung wird ebenso in der Forschung von Magrabi et al. [2018] geteilt.
- *Fehlerreduktion*: Spracherkennungssysteme, die gesprochene Sprache in Text umwandeln, tragen dazu bei, typische Fehler der manuellen Dateneingabe zu vermeiden. Dies ist besonders kritisch laut Kim et al. [2021b], da selbst geringfügige Fehler in medizinischen Dokumentationen schwerwiegende Konsequenzen nach sich ziehen können. Eine Studie von Zuchowski und Göller [2022] ergab, dass die Fehlerquote bei Verwendung von Spracherkennungssystemen bei nur 0,15 Fehlern pro Zeile liegt, verglichen mit 0,3 Fehlern pro Zeile bei manueller Eingabe.
- *Verbesserte Befundqualität und -vollständigkeit*: Ärzt:innen können mithilfe der Spracherkennung detaillierte und umfassende Befunde erstellen, ohne sich über Schreibgeschwindigkeit oder -aufwand Gedanken machen zu müssen. Das Paper von Blackley et al. [2020] zeigte, dass diktierete Berichte umfassender waren und mehr relevante Informationen enthielten. Ähnliche Ergebnisse zeigt eine Studie von Magrabi et al. [2018], mit dem Ergebnis, dass Befunde länger und von besserer Qualität in Bezug auf Klarheit und Vollständigkeit waren.
- *Verbesserte interdisziplinäre Kommunikation*: Die sofortige Verfügbarkeit digital erstellter Dokumente erleichtert den Austausch zwischen verschiedenen Fachbereichen, verbessert die interdisziplinäre Kommunikation und fördert eine koordinierte Patientenversorgung. Laut einer Studie von Magrabi et al. [2018] mit dem Titel "Evaluating the usability of speech recognition to create clinical documentation using a commercial electronic health record" führt der Einsatz von Spracherkennungstechnologien zu einer Vereinheitlichung des klinischen Workflows und einer verbesserten Kommunikation zwischen den Abteilungen. Weiters erleichtern diese Technologien die Durchführung von Telemedizin und Fernkonsultationen, wie in den Studien von Kumar [2024] und Kim et al. [2021b] dargestellt wird.
- *Reduzierter Papierverbrauch*: Die digitale Erfassung von Daten trägt zur Reduktion des Papierverbrauchs bei. Dies senkt nicht nur die Betriebskosten, sondern

unterstützt auch eine Verringerung des ökologischen Fußabdrucks im Gesundheitswesen. Diese Beobachtung wird durch die Studie von Kumar [2024] bestätigt.

Herausforderungen

- *Sicherheits- und Genauigkeitsbedenken:* Trotz der Fortschritte in der Anwendung von Künstlicher Intelligenz im Gesundheitswesen bestehen laut Latif et al. [2020] Bedenken hinsichtlich der Sicherheit. Die Modelle sind anfällig für "adversarial attacks", die Fehlentscheidungen provozieren können.
- *Datenmanagement und -integration:* Im Gesundheitssektor erschwert durch die uneinheitliche Datenerfassung laut Deepa und Khilar [2022] nicht nur den Vergleich von Ergebnissen, sondern auch die Entwicklung präziser diagnostischer Methoden. Zudem beeinträchtigt die mangelnde Datenkompatibilität zwischen medizinischen Geräten laut Latif et al. [2020] den Einsatz moderner Technologien und wirkt sich negativ auf die Patientenversorgung aus. Daher ist eine verbesserte Interoperabilität entscheidend, um die Effizienz von Spracherkennungstechnologien voll ausschöpfen zu können.
- *Anpassungs- und Nutzungsherausforderungen:* Die Einarbeitung in Spracherkennungssysteme kann im Vergleich zur Verwendung traditioneller Eingabemethoden wie Tastatur und Maus zeitaufwendig und frustrierend sein. Dies umfasst auch laut Kumar [2024] Herausforderungen wie Fehlinterpretationen, Transkriptionsfehler und den Umgang mit medizinischer Fachterminologie. Sprachvariationen, insbesondere bei mehrsprachigen Benutzer:innen, stellen dabei eine weitere Herausforderung für die Genauigkeit von Spracherkennungssystemen dar.
- *Technische Einschränkungen und Benutzerpräferenzen:* Die Abhängigkeit von einer stabilen Netzwerkverbindung für cloud-basierte Programme wie "Dragon Medical SpeechKit" kann es in Bereichen mit schlechter Internetanbindung zu Unterbrechungen der Spracherkennung führen. Weiters bevorzugten Studienteilnehmer laut Blackley et al. [2020] trotz der verfügbaren Technologie, teilweise das Tippen und nutzten nur begrenzt verfügbare Shortcuts, was die potenzielle Effizienz von Spracherkennung mindert. Zusätzlich schätzten Kliniker laut Zuchowski und Göller [2022] die Fehlerquote von Spracherkennungssoftware höher ein als bei manueller Eingabe.
- *Ethische Bedenken:* Vorurteile und mangelhafte Trainingsdaten können laut Brandon [2021] dazu führen, dass Spracherkennungssysteme ungerechte oder diskrimi-

nierende Entscheidungen treffen. Es ist daher entscheidend klarzustellen, wer die Verantwortung für diese Probleme trägt und welche Schritte unternommen werden können, um derartige Verzerrungen zu minimieren. Dies erfordert die Implementierung systematischer Überprüfungen und Audits, die nicht nur die Datenqualität gewährleisten, sondern auch potenzielle Vorurteile aufdecken.

3.4.1 Datenschutzaspekte

Bei der Verarbeitung von Patientendaten wie auch Sprachaufnahmen müssen laut Gesetz bestimmte Datenschutzaspekte gegeben sein. In der österreichischen Gesetzgebung ist das Datenschutzgesetz (DSG) [RIS, 2023] sowie die Europäische Datenschutzgrundverordnung (DSGVO) [EU-Parlament, 2016] zu berücksichtigen. In den Gesetzen ist verankert, dass die folgenden Aspekte gegeben sein müssen.

- *Einwilligung*: Vor der Verarbeitung von Sprachdaten muss die Einwilligung der Patient:innen und des medizinischen Personals oder der Ärzt:innen eingeholt werden. Diese Einwilligung muss spezifisch, informiert und eindeutig formuliert sein.
- *Datensicherheit*: Gemäß DSG und DSGVO müssen technische und organisatorische Maßnahmen getroffen werden, um ein hohes Sicherheitsniveau zu gewährleisten. Dazu gehört der Schutz vor unbefugtem Zugriff, sowie vor Verlust, Zerstörung oder Schädigung der Daten.
- *Speicherbegrenzung*: Die Speicherung von Sprachdaten ist auf das erforderliche Maß zu beschränken. Wenn sie nicht mehr benötigt werden, sind sie zu löschen oder gemäß den gesetzlichen Bestimmungen zu anonymisieren.
- *Zugriffsbeschränkung*: Der Zugriff muss streng reglementiert sein und nur Personen gewährt werden, die aufgrund ihrer Arbeitsaufgaben Zugang benötigen. Die Zugriffsprotokolle sind zu sichern.
- *Transparenz*: Organisationen sind dazu verpflichtet, transparent über die Verwendung von Sprachdaten zu informieren.

3.5 Unternehmen

Es gibt laut Chen und Decary [2020] drei Kategorien von Unternehmen, die Produkte und Dienstleistungen für die Entwicklung von Spracherkennungssystemen im Gesundheitswesen anbieten.

Die erste Kategorie besteht laut Eggert und Kreuzer [2022] aus Anbietern elektronischer Patientenakten, die damit begonnen haben, ihre eigenen KI-Funktionen in ihre Systeme zu integrieren, indem sie maschinelles Lernen, Sprachdiktate und Sprachprozesse zur Unterstützung klinischer Entscheidungen und Arbeitsabläufe einsetzen. Diese Unternehmen haben beispielsweise Anwendungen wie Spracherkennung und Systeme zur Unterstützung klinischer Entscheidungen entwickelt.

Die zweite Kategorie umfasst große Technologieunternehmen [Vogel und Varoß, 2024] wie Google, Microsoft, Amazon, Apple, Nuance und IBM, die KI-Cloud-Systeme für Gesundheitsorganisationen bereitstellen. Diese Unternehmen haben Systeme zur Bilderkennung, Spracherkennung und prädiktiven Analyse sowie KI-Assistenten wie "Alexa", "Siri", "Cortana" und "Google Assistant" entwickelt. Speziell für das Gesundheitswesen gibt es Systeme wie "IBM Watson" [Bajpai und Sharda, 2022], das Diagnose- und Behandlungsempfehlungen bei Krebs geben kann. Darüber hinaus bietet Nuance Spracherkennungstechnologien an, die in elektronische Patientenakten integriert werden können, um die klinische Dokumentation und Dateneingabe zu unterstützen. Weiters hat Google DeepMind laut Yim et al. [2020] ein System entwickelt, das Diagnose- und Behandlungsempfehlungen für mehr als fünfzig Augenkrankheiten gibt.

Die dritte Kategorie umfasst Unternehmen, die sich auf die Entwicklung künstlicher Intelligenz für das Gesundheitswesen spezialisiert haben. Zu ihren Innovationen gehören Chatbots für Patienten, wie beispielsweise "Woebot" [Darcy et al., 2022] - ein Chatbot für psychische Gesundheit, der Patienten bei der Vorbereitung auf chirurgische Eingriffe unterstützt. Diese Unternehmen entwickeln aber auch Spracherkennungstools für die klinische Dokumentation wie "Voicedocs" [Voicedocs, 2022] eine, Spracherkennung für Diktat und Transkription im Gesundheitswesen.

Kapitel 4

Methoden

4.1 Literaturrecherche

Um den theoretischen Teil der Forschungsfragen umfassend zu beantworten und einen gründlichen Überblick über das Thema Spracherkennung in der Medizin zu erhalten, wurde eine ausgedehnte Internet- und Literaturrecherche durchgeführt. Ziel dieser Recherche war es, den aktuellen Stand der Technik sowie die neuesten Entwicklungen in diesem Bereich zu erfassen. Zunächst erfolgte eine umfangreiche Internetrecherche, um aktuelle Informationen und Trends zu identifizieren. Parallel dazu wurde eine systematische Literatursuche in verschiedenen wissenschaftlichen Datenbanken durchgeführt. Zu den herangezogenen Datenbanken zählten Plattformen wie Google Scholar, PubMed, ScienceDirect, SpringerLink, die Bibliothek der FH Joanneum, und das "Institute of Electrical and Electronic Engineers (IEEE)". Es wurden gezielt Schlagworte wie "Spracherkennung", "speech recognition", "speech recognition in medicine", "speech recognition in healthcare", "Deep Learning methods in speech recognition", "Machine Learning", "speech recognition healthcare history" und "future of speech recognition" verwendet. Die Auswahl und Bewertung der relevanten wissenschaftlichen Arbeiten erfolgte nach den PRISMA-Richtlinien. Zusätzlich wurde das Internet gezielt nach Begriffen rund um die Anwendung von Spracherkennungstechnologien im Gesundheitswesen durchsucht. Dazu zählten spezifische Suchanfragen nach "databanks speech recognition healthcare", "speech models in speech recognition", „Dragon Medical One“, "Dragon Medical Speechkit" sowie nach Bewertungen und Konkurrenzprodukten dieses Systems. Auch wurde nach spezifischen Einsatzmöglichkeiten der Spracherkennung in verschiedenen medizinischen Fachbereichen gesucht, um ein möglichst breites Verständnis der Einsatzpotenziale und Herausforderungen dieser Technologie zu erlangen. Die folgende Grafik zeigt den Verlauf der Literaturauswahl nach den "PRISMA-Richtlinien". Der Prozess ist in vier Phasen aufgeteilt: Identifikation, Vorauswahl, Eignung und letztendlich Einschluss.

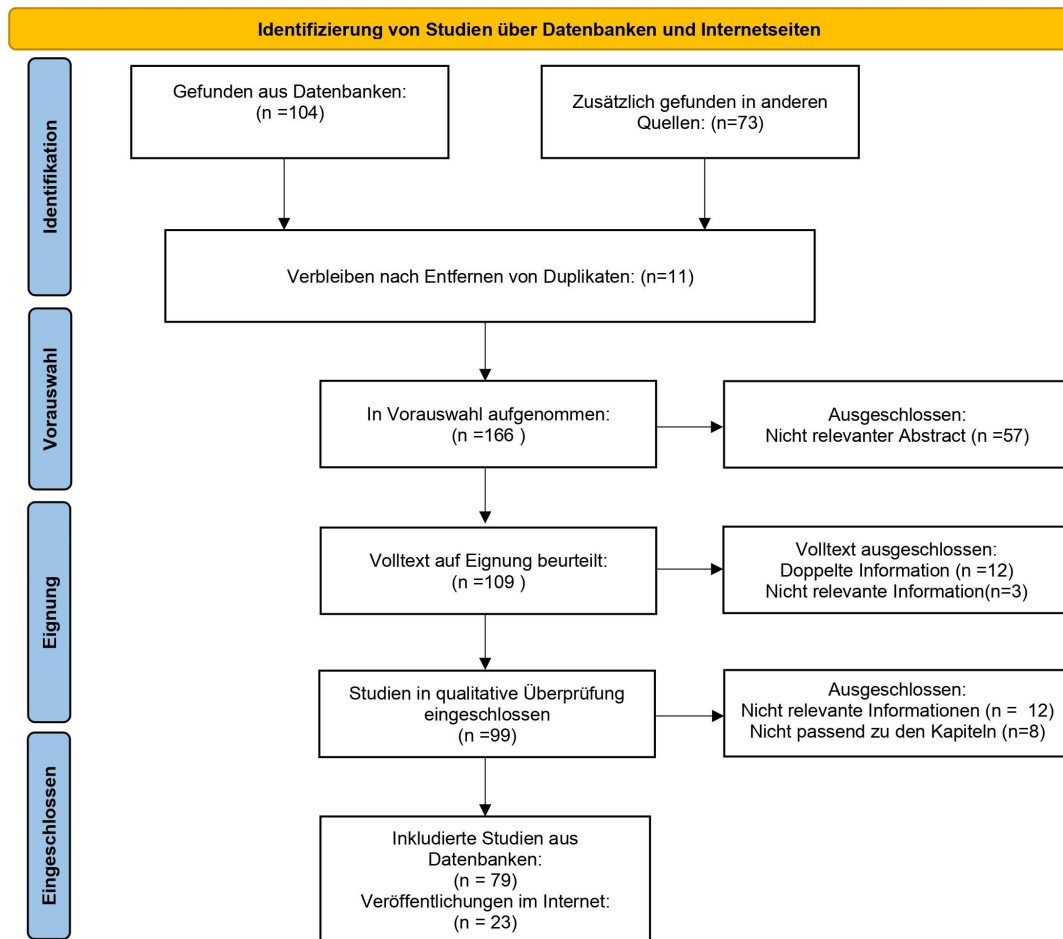


Abbildung 4.1: Auswahlprozess der Literatur nach den PRISMA Richtlinien. Aufgeteilt in Literatur von Datenbanken und Internetwebseiten. Die Literatursuche wurde innerhalb von vier Phasen durchgeführt: erste Identifikation, erste Vorauswahl, mögliche Eignung und letztendlich Einschluss in die Arbeit.

4.2 Interview

Um die Herausforderungen bei der Implementierung von "Dragon Medical SpeechKit" in den klinischen Workflow der Barmherzigen Brüder Graz umfassend zu verstehen, wurde ein Interview mit einem Mitglied des IT-Business Management durchgeführt. Er ist zuständig für die Betreuung der Ausrollung vom "Dragon Medical SpeechKit" in ganz Österreich. Die Vorbereitung für das Interview umfasste die Erstellung von sieben Fragen, die darauf abzielten, ein tiefgehendes Verständnis der technischen Aspekte der Einführung zu erlangen. Zusätzlich wurden weitere Fragen formuliert, um die spezifischen Herausforderungen und potenziellen Probleme zu identifizieren, die während der Implementierungsphase auftraten.

Das Interview fand in einer Microsoft Teams Besprechung statt. Nach der automatischen Transkription durch die Software nahm die Autorin eine Überprüfung des Transkripts vor, um kleinere Fehler zu korrigieren und die Genauigkeit der wiedergegebenen Informationen sicherzustellen.

4.3 Beobachtungsanalyse

Um spezifische Herausforderungen und die Nutzerzufriedenheit bei der Integration von Spracherkennungstechnologien in den klinischen Workflow zu untersuchen, wurde eine Beobachtungsanalyse auf zwei Stationen für Innere Medizin durchgeführt. Im Zeitraum vom 25.3.2024 bis zum 5.5.2024 nahmen 26 Testpersonen (zwanzig weiblich, sechs männlich) diplomierte Gesundheits- und Krankenpfleger:innen (DGKP) an der Studie teil. Die Mehrheit dieser Personen verfügte bereits über Erfahrungen mit "Dragon Medical One" und "Dragon Medical Speechkit". Die Beobachtungen erfolgten offen und fanden täglich zwischen 11 und 13 Uhr, im Anschluss an die Visite, statt. In diesem Zeitfenster wurden die Teilnehmenden gezielt nach ihren Erfahrungen und eventuell auftretende Problemen mit der Software befragt. Ziel war es, ein klares Bild davon zu bekommen, wie gut die Technologie in den alltäglichen klinischen Betrieb integriert werden kann und welche Störfaktoren möglicherweise auftreten. Die während der Studie gesammelten Daten wurden in einer strukturiert angelegten Excel-Datei festgehalten. Die kategorische Strukturierung dieser Datei umfasste mehrere Abschnitte: Namen der Teilnehmer:innen, Benutzernamen, Beschreibungen der aufgetretenen Fehler, Tastenfunktionen und Computernamen, das jeweilige Datum und zusätzliche Notizen. Ein zentrales Anliegen dieser Analyse war die Einhaltung ethischer Standards und der Schutz der Privatsphäre der Teilnehmenden. Vor

Beginn der Beobachtungsanalyse wurde von allen Teilnehmenden eine Einwilligung eingeholt, nachdem sie über das Ziel der Studie informiert wurden.

4.4 Fragebogen

Um die Forschungsfragen zu beantworten, wurde ein Fragebogen entwickelt. Der Fragebogen umfasste 24 Fragen, die auf die Bewertung der Benutzerfreundlichkeit, Genauigkeit, Vollständigkeit der Spracherkennung, Effizienz des Systems, Qualität der Schulungen und die allgemeine Zufriedenheit mit Spracherkennungssystemen abzielten. Bei den Fragen wurde mithilfe von künstlicher Intelligenz auf eine klare und direkte Fragestellung geachtet, um Missverständnisse zu vermeiden. Die Mehrheit der Fragen wurde mittels der Likert-Skala bewertet, die von eins (trifft zu) bis fünf (trifft nicht zu) reichte. Dies ermöglichte eine standardisierte Bewertung der Antworten. Zusätzlich gab es Freitextfragen, die den Teilnehmenden die Möglichkeit gaben, spezifische Funktionen oder Aspekte detaillierter zu kommentieren. Der Fragebogen wurde mithilfe von Microsoft Forms erstellt, einer Plattform, die sich für das Erstellen von Quizze, Formularen und Umfragen eignet. Die Zielgruppe für die Umfrage umfasste diplomierte Gesundheits- und Krankenpfleger:innen verschiedener Stationen, die während des Ausrollungsprozesses als Teststationen fungierten. Die Datenerhebung fand von Ende April bis Ende Mai 2024 statt, wobei alle Teilnehmer vor der Beantwortung des Fragebogens ihre informierte Einwilligung gaben. Vor dem offiziellen Start der Befragung wurde eine Pilotphase durchgeführt. Während dieser Phase wurden zwei Stationsleitungen gebeten, den Fragebogen hinsichtlich der Verständlichkeit der Fragen und der technischen Funktionalität zu bewerten. Dieser Schritt diente dazu, eventuelle Mängel im Fragebogen frühzeitig zu identifizieren und zu korrigieren. Großer Wert wurde auf den Datenschutz und die Anonymität der Teilnehmenden gelegt. Es wurde sichergestellt, dass aus den Antworten nicht auf die Identität der Befragten zurückgeschlossen werden kann. Alle gesammelten Daten wurden anonym verarbeitet und nach Abschluss der Studie vernichtet, um die Vertraulichkeit zu wahren. Für die Auswertung der Umfragedaten wurden die Analysewerkzeuge von Microsoft Forms genutzt. Diese Werkzeuge ermöglichten eine automatische Auswertung der Antworten für jede Frage sowie eine tabellarische Darstellung der Altersdemografie der Stichprobe.

Kapitel 5

Resultate

5.1 Dragon Medical SpeechKit

Das Unternehmen Nuance wurde 1975 gegründet und brachte 1982 seine erste Spracherkennungstechnologie auf den Markt. Es ist laut Onitilo et al. [2023] in den Bereichen Künstlicher Intelligenz und biometrischer Sicherheit aktiv und unterstützt weltweit etwa zehntausend Gesundheitsorganisationen mit seinen Produkten. Diese umfassen eine cloud-basierte Lösung, die sich durch den Einsatz von KI an die natürlichen Arbeitsabläufe medizinischen Personals anpasst. Zu den integrierten Funktionen zählen laut Nuance [2024] die Erstellung persönlicher Terminologie und Vorlagen, die auf verschiedenen Geräten verwendet werden können. Das System verbessert sich kontinuierlich durch automatische Prozesse wie Mikrofonkalibrierung und Akzentanpassung.

Zusätzlich zu "Dragon Medical One", das seit 2022 im ärztlichen Bereich im Haus der Barmherzigen Brüder Graz eingesetzt wird, entwickelte Nuance das "Dragon Medical SpeechKit". Diese cloudbasierte Spracherkennung ist Teil der "Nuance Healthcare Development Platform" und ermöglicht die Spracherkennung sowohl in Desktop-Browsern als auch auf mobilen Geräten. Die Vorteile des "SpeechKits" liegen in der schnellen Integration und der minimalen Belastung durch das "Null-Design-Prinzip", das keine Belastung des Clients, keinen Implementierungsaufwand und keinen Integrationsaufwand erfordert. Das SpeechKit besteht aus vier Hauptkomponenten:

- *SpeechMagic*: Ermöglicht die nahtlose Integration von Spracherkennungsanwendungen, verantwortlich für digitales Diktat, Front-End- und Back-End-Spracherkennung.
- *Dragon Medical SpeechKit*: Set von cloud- und netzwerkbasierenden Webdiensten, das sich durch das Prinzip der "Null-Belastung des Clients, Null-Implementierungsaufwand und Null-Integrationsaufwand" auszeichnet.

- *Dragon Medical Workflow Edition*: Ein Set für digitales Diktat, Spracherkennung und Workflow-Integration im Krankenhausinformationssystem (KIS).
- *Nuance Healthcare Developer Programm*: Ermöglicht die Integration von Sprachfunktionen in bestehende klinische Applikationen und Erstellung von spezifischem Vokabular für Krankenhäuser oder Abteilungen.

Die Studie [Saxena et al., 2018] berichtete, dass 81,8 Prozent der Nutzer vom "Dragon Medical SpeechKit" eine Verbesserung sowohl in der Länge als auch im Detailgrad ihrer Dokumentation feststellten. Weiters konnte das "Vassar Brothers Medical Center" die Transkriptionskosten um 81,3 Prozent senken, was Einsparungen von 1,37 Millionen US-Dollar entsprach. Eine Vergleichsstudie am Universitätsklinikum Düsseldorf zeigte laut Vogel et al. [2015] ähnliche Ergebnisse: Die Dokumentationsgeschwindigkeit mit Spracherkennung erreichte 217 Zeichen pro Minute im Vergleich zu 173 Zeichen ohne Spracherkennung, was die Dokumentationszeit um 26 Prozent reduzierte und gleichzeitig den Detailgrad der Dokumentation um 82 Prozent steigerte.

Das System unterstützt 23 Sprachen, verarbeitet Audio in Echtzeit, doch Qualität der Spracherkennung ist stark von der Netzwerkverbindung abhängig. Technologien wie Audiokompression und die SpeakAhead-Technologie verbessern die Performance, indem sie es Nutzer:innen ermöglichen, trotz Netzwerklatenz weiterzudiktieren, während die Texterkennung asynchron verarbeitet wird. Zudem entspricht die Verschlüsselung den europäischen Sicherheits- und Datenschutzbestimmungen. Die Lizenzkosten betragen 35 Euro pro Monat je Benutzer.

Insgesamt zielt das "Dragon Medical SpeechKit" darauf ab, Spracherkennungstechnologie nahtlos und effizient in klinische Anwendungen zu integrieren, um eine vollständigere und zeitsparende Patientendokumentation zu ermöglichen.

Weiters folgt ein Vergleich von vier Spracherkennungssysteme für das Gesundheitswesen: "Dragon Medical SpeechKit" [Nuance, 2016], "MModal Fluency Direct" [Byrne, 2023] [3M, 2024], "Amazon Transcribe Medical" AWS [2024], und "Olympus Dictation" [OM Digital Solutions, 2024], diese Systeme wurden für die Unterstützung im medizinischen Alltag entwickelt. In der unten stehenden Tabelle werden sie anhand der folgenden Kategorien beurteilt: allgemeines medizinisches Vokabular, wie "Patientenaufnahme", "Befund" und weitere, medizinisches Vokabular speziell für bestimmte Fachgebiete, die Möglichkeit der Erstellung von personalisierten Vokabular (oft genutzte Abkürzungen), die Integration in bestehende Krankenhaussysteme, gegebene Datensicherheit laut DSGVO, intuitive Benutzeroberfläche und Skalierbarkeit, die Fähigkeit eines Systems mitzuwachsen.

Tabelle 5.1: Vergleich von vier Spracherkennungstechnologien, die zur Unterstützung des medizinischen Personals in Krankenhäusern entwickelt wurden.

Kriterium	Dragon Medical SpeechKit	MModal Fluency Direct	Amazon Transcribe Medical	Olympus Dictation
Medizinisches Vokabular	+	+	+	+
Spezifisches Vokabular für med. Fachbegriffe	+	+	+	-
Personalisiertes Vokabular	+	-	-	-
Integration in bestehende Systeme	+	+	+	-
Datensicherheit	+	+	+	+
Benutzerfreundlichkeit	+	+	+	+
Skalierbarkeit	+	+	+	-

5.1.1 Einführung in den klinischen Workflow

Laut dem Interviewpartner, der schon bei der erfolgreichen Implementierung und dem Rollout des "Dragon Medical One" (DMO) bei den Ärzt:innen im Haus der Barmherzigen Brüder Graz, mitwirkte, wurde beschlossen, auch das Pflegepersonal mit einem Produkt von Nuance auszustatten. Die Entscheidung fiel dabei auf das "Dragon Medical Speechkit" (DMSK), das im weiteren Verlauf in den klinischen Workflow integriert werden sollte. Bereits im Jahr 2021 wurde vom IT-Business Management die Entscheidung für das "Dragon Medical Speechkit" getroffen und ein Nutzungsvertrag über drei Jahre abgeschlossen. Die Implementierung der DMSK-Technologie begann mit einer Auswahl der Technologie und einer Überprüfung der Datenschutzrichtlinien, gefolgt von der Einführung spezifischer Sicherheitsprotokolle. Beide Lösungen sind laut dem Interviewpartner, sind cloudbasierte Dienste, deren Datenübertragung entsprechend verschlüsselt erfolgt (HTTPS). Um die neue Technologie reibungslos in den klinischen Alltag zu integrieren, wurden die ersten Implementierungsschritte 2023 auf einer Station für Innere Medizin durchgeführt. Diese Testphase sollte die Einführung erleichtern und mögliche Probleme frühzeitig identifizieren. Für die Testphase wurde die notwendige Hardware und zusätzliche Software, einschließlich Mikrofone (SpeechMike Premium) und Lizenzen für das "Nuance SpeechKit" angeschafft. Geplante Treffen und Schulungen für die diplomierten Gesundheits- und Krankenpflegepersonen pro Abteilung waren ebenfalls Teil der Vorbereitungen. Nach der erfolgreichen Testphase wurde die Implementierung für Frühjahr 2024 angesetzt. Der Rollout für die Teststation war ursprünglich für den 22.03.2024 geplant, musste jedoch aufgrund von Lieferverzögerungen bei den Lizenzen auf den 28.03.2024 verschoben werden. Seit dem 03.04.2024 werden tägliche Audits durch die IT-Abteilung durchgeführt, um die Nutzung und Integration der neuen Technologie zu überwachen. Die Spracherkennung wurde von fast allen Mitarbeiter:innen zur Dokumentation genutzt, wobei besonders das männliche Personal die Technologie intensiver nutzte. Ein weiterer Vorteil der neuen Lösung ist laut dem Interviewpartner die problemlose Integration in den vorhandenen Programmworkflow, da das Programm wie eine Tastatur funktioniert und den erkannten Text an der Cursor-

Position einfügt. Um einen kontinuierlichen Austausch zwischen den Verantwortlichen zu gewährleisten, fanden wöchentliche Meetings mit der IT-Leitung, der IT-Projektleitung, der Business-Informatik und der Pflegeinformatik statt. Nachdem die erste Teststation eine regelmäßige Nutzung ohne Fehler und Probleme zeigte, wurde beschlossen auf weitere Stationen auszurollen. Während des Rollouts traten jedoch einige technische Probleme auf. Die Nutzernamen im Active Directory, die KIS-Nutzernamen und die Lizenznamen stimmten nicht überein, was zu Login-Fehlern führte. Dieser Fehler wurde rasch behoben, und die Technologie konnte weiter ausgerollt werden. Die nächste Station, die in den Rollout einbezogen wurde, war Ende April 2024 eine weitere Abteilung für Innere Medizin. Diese Station hatte zuvor keinerlei Erfahrung mit Spracherkennung oder dem "Dragon Medical Speechkit". Die Einführung auf der zweiten Station wurde durch regelmäßige Schulungen und tägliche Überprüfungen der Nutzungsdaten begleitet. Die Nutzerakzeptanz wurde durch einen Fragebogen und den tatsächlichen Nutzungsdaten des Systems beurteilt. Dabei zeigte sich, dass einige Benutzer zusätzliche Unterstützung benötigten, um die Technologie effizient zu nutzen. Nach den erfolgreichen Einführungen auf den Stationen verlief der Rollout auf weiteren Stationen ähnlich strukturiert. Auf diesen Stationen wurde die Hardware verteilt und Schulungen mit der Abteilungsleitung sowie den DGKPs durchgeführt. Anschließend fanden weitere Schulungen statt, bis alle DGKPs auf den Stationen geschult waren. Diese systematische Vorgehensweise stellte sicher, dass die neue Spracherkennungstechnologie effektiv in den klinischen Alltag integriert wurde. Die Implementierung von "Dragon Medical SpeechKit" hat gezeigt, dass sorgfältige Planung, regelmäßige Schulungen und kontinuierliche Überwachung entscheidend für die Akzeptanz und den Erfolg neuer Technologien im klinischen Umfeld sind.

5.2 Datenanalyse Beobachtungsanalyse

Zu Beginn der Beobachtungsanalyse meldeten neun der 26 Testpersonen (34,6 %), dass sie beim Versuch zu diktieren, eine Fehlermeldung erhielten, die auf das Fehlen einer Lizenz hinwies. Dies wurde weitergehend untersucht. Während der direkten Beobachtungsphase stellte sich heraus, dass das Kernproblem im Nutzerverhalten lag. 85 % der Testpersonen setzte sich zu beliebigen Computern und meldeten sich nicht mit ihren eigenen Benutzerdaten an, sondern nutzten stattdessen die bereits angemeldeten Konten weiter. Diese Praxis führte zu einer Verwirrung bei der Zuweisung von Lizenzen, da eigentlich lizenzierte Benutzer aufgrund der Verwendung fremder Konten als unlicenziert wahrgenommen wurden. Eine besondere Herausforderung ergab sich aus der Art und Weise, wie die Lizenzen von Nuance verteilt wurden. Da die Lizenzen auf die gesamte Steiermark ausgedehnt

wurden, entstanden Schwierigkeit, wenn Personen aus verschiedener Krankenhäuser, denselben Benutznamen hatten. Diese Situation machte es unmöglich, ihnen eine Lizenz unter ihrem Benutzernamen zuzuweisen. Folglich mussten die Benutzernamen von sechs Testpersonen geändert werden. Zudem kam es zu einer Verwechslung im Active Directory, da zwei Personen denselben Namen und Benutzernamen hatten, was dazu führte, dass die IT die falsche Person zur Testgruppe hinzufügte. Ein ergonomisches Problem mit dem Mikrofon waren die Tastenanordnungen, die dazu führte, dass der Mauszeiger bei natürlicher Handhaltung ungewollte Bewegungen machte. Zusätzlich wurde bei 27 Prozent der Laptops festgestellt, dass eine unzureichende Internetverbindung zu Netzwerkfehlern führte. Ein weiteres Problem, in der zweiten Teststation war, dass ein Programm auf den Laptops der Station fehlte, um die Aufnahmetaste des Mikrofons betätigen zu können, was zu sinkender Akzeptanz des gesamten Programms führte.

Auf der ersten Station wurde die SE gut angenommen und wird nun täglich von den Mitarbeitern benutzt, bei der zweiten Station die ausgerollt wurde, die bis jetzt noch nie mit dieser Spracherkennungssoftware gearbeitet haben, gab es einige Hürden bis die Mitarbeitenden zur regelmäßigen Nutzung motiviert werden konnten. Bei der dritten Station zeigte sich schon am Anfang, dass die Nutzung mäßig ausfallen wird, da sich schon Vorurteile vom Spracherkennungssystem "Dragon Medical One", das von den Ärzten benutzt wird, übertragen haben. Das anzuschließende Kabel und das frei am Tisch liegende Mikrofon wurde als störend empfunden. Weiters war das freie Sprechen im Pflegestützpunkt mit Hemmungen verbunden.

Insgesamt wurde "Dragon Medical Speechkit" aber gut angenommen, besonders die Textbausteine überzeugten viele DGKPS, da ein Stichwort und der Befehl "einfügen" genügt. Die Nutzer können den Textbaustein weiter personalisieren oder die Dokumentation einfach speichern. Besonders bei Stationen für Innere Medizin, wo oft Endoskopie-Patienten Übernahmen vorkommen, erwiesen sich solche Textbausteine als bedeutend.

Tabelle 5.2: Darstellung der häufigsten Beobachtungen der durchgeführten Analyse, die auf den Teststationen durchgeführt wurde.

Kategorie	Hauptbeobachtung	Empfohlene Maßnahme
Lizenzmanagement	Lizenzfehler aufgrund fremder Benutzerkonten	Striktere Anmeldeverfahren und verbesserte Lizenzverwaltung
Technische Probleme	Kompatibilitätsprobleme mit Citrix und Philips-Mikrofonen	Technische Überprüfung und Anpassung der Systeme
Ergonomische Herausforderungen	Ungewollte Cursorbewegungen durch Mikrofonbedienung	Ergonomische Überarbeitung der Hardware
Netzwerkprobleme	Unzureichende Internetverbindung führte zu Netzwerkfehlern	Verbesserung der Netzwerkinfrastruktur
Softwareanforderungen	Fehlende Software für notwendige Funktionen	Sicherstellung der vollständigen Softwareinstallation
Akzeptanz der Technologie	Vorurteile und Anlaufschwierigkeiten an neuen Stationen	Gezielte Schulungen und Informationsvermittlung



Abbildung 5.1: Darstellung eines Philips SpeechMike mit integrierter Maus (grün markiert), bei dem unbeabsichtigte Berührungen des Feldes zu unerwünschten Cursorbewegungen führen können. [Philips, 2024]

Die Analyse dieser Beobachtungen zeigt, dass eine Reihe von Faktoren – sowohl menschliches Verhalten als auch technische Einschränkungen – die Effektivität der Implementierung von Spracherkennungstechnologien beeinträchtigen. Um diese Herausforderungen zu überwinden, wird empfohlen:

- Einführung strengerer Anmeldeverfahren und ausführliche Schulungen für das Personal, um das Bewusstsein für die Bedeutung der individuellen Anmeldung zu schärfen.
- Überprüfung und Anpassung der Lizenzverwaltungspraktiken, um Probleme mit doppelten Benutzernamen zu vermeiden.
- Technische Überprüfung und Anpassung der Hardwarekonfigurationen, um Kompatibilitäts- und Ergonomieprobleme zu minimieren.

Durch die Umsetzung dieser Empfehlungen kann die Zuverlässigkeit und Benutzerfreundlichkeit von Spracherkennungstechnologien im klinischen Umfeld wesentlich verbessert werden. Insgesamt zeigten männliches und jüngeres Personal die größte Benutzerakzeptanz.

5.3 Datenanalyse Fragebogen

Im Rahmen dieser Studie wurde ein Fragebogen von 32 DGKPs der Barmherzigen Brüder Graz beantwortet. Die Altersverteilung, die mögliche Einflüsse auf das Nutzerverhalten haben kann, war wie folgt aufgeteilt: Kein Teilnehmer war zwischen 18 und 20 Jahren, 13 Teilnehmer:innen waren zwischen 21 und 29 Jahren, zehn zwischen 30 und 39 Jahren, sechs zwischen 40 und 49 Jahren und drei zwischen 50 und 59 Jahren.

Benutzerfreundlichkeit

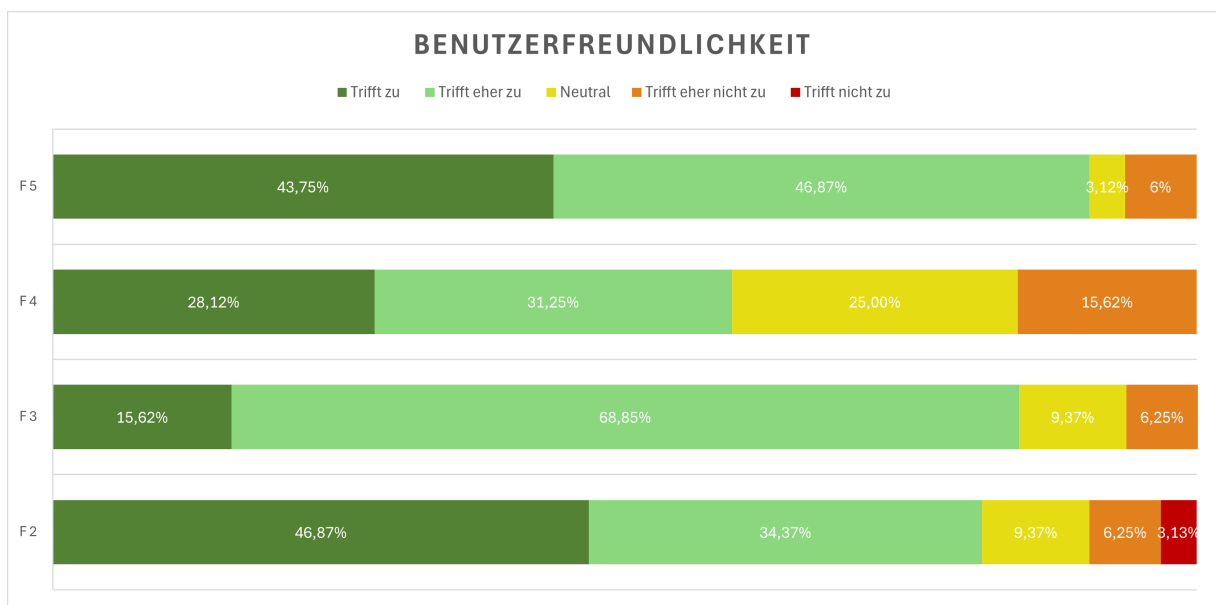


Abbildung 5.2: Auswertung der Fragen des Fragebogens mit den Themen Gewöhnung, intuitive Benutzeroberfläche und Verbesserung der Arbeitsweise. F2: Konnten Sie sich leicht an die grundlegenden Funktionen der Spracherkennung gewöhnen?, F3: Unterstützt die Benutzeroberfläche eine effektive Arbeitsumgebung?, F4: Hat die Verwendung der Spracherkennung zu Verbesserungen in Ihrem Arbeitsablauf geführt?, F5: Empfinden Sie die Handhabung des Systems als intuitiv?

Bei der Kategorie Benutzerfreundlichkeit ging eine Frage um, wie leicht sich die Teilnehmer an die grundlegenden Funktionen von Dragon gewöhnen konnten, dabei lag die durchschnittliche Bewertung bei 1,84 (SD= 1,05). Dies deutet darauf hin, dass die Mehrheit die Gewöhnung als einfach empfand, wobei 15 Personen "stimme zu" und 11 "stimme eher zu" wählten. Dennoch traten anfänglich technische Probleme auf, insbesondere mit dem Mikrofon-Button, der aufgrund fehlender Treiber nicht betätigt werden konnte. Benutzer mussten als Workaround manuell auf den Bildschirm klicken, was nach der Treiberinstallation teilweise zu Verwirrung führte.

In Bezug auf die Unterstützung der Benutzeroberfläche von Dragon für eine effektive Befunderfassung ergab sich eine durchschnittliche Bewertung von 2,06 (SD= 0,71). Fünf Personen stimmten voll zu, während 22 "stimme eher zu" wählten und drei neutral blieben. Viele Teilnehmer äußerten sich positiv, da sie das System wie gewohnt weiter nutzen konnten. Gleichzeitig zeigte sich jedoch, dass nicht alle Teilnehmer der Meinung waren, dass die Verwendung von Dragon zu einer Verbesserung in ihrer Arbeitsweise geführt hat. Dies spiegelt sich in der durchschnittlichen Bewertung von 2,28 (SD= 1,05) wider, wobei neun Teilnehmer voll zustimmten, zehn eher zustimmten, acht neutral blieben und fünf eher nicht zustimmten. Während einige die Textbausteine als nützlich empfanden, wurde das zusätzliche Kabel von anderen als störend empfunden.

Die Frage, ob die Handhabung von Dragon als intuitiv empfunden wurde, ergab eine durchschnittliche Bewertung von 1,72 (SD= 0,81). Hier stimmten 14 Teilnehmer voll zu, 15 eher zu, eine Person blieb neutral und zwei stimmten eher nicht zu. Die Teilnehmer fanden es leicht, sich daran zu gewöhnen, eine Taste mit einem kleinen roten Punkt zu drücken, um die Spracherkennung zu starten und zu stoppen. Allerdings gab es anfängliche Schwierigkeiten, da manche Nutzer vergaßen, die Aufnahme zu stoppen oder dachten, sie müssten den Button gedrückt halten.

Zusammenfassend lässt sich sagen, dass die Mehrheit der Teilnehmer mit der Gewöhnung an die grundlegenden Funktionen von Dragon zufrieden war, obwohl es anfängliche technische Schwierigkeiten gab. Die Benutzeroberfläche wurde überwiegend positiv bewertet, was die Unterstützung einer effektiven Befunderfassung angeht. Die Meinungen zur Verbesserung der Arbeitsweise durch Dragon waren gemischt, wobei einige die neuen Funktionen als hilfreich empfanden, während andere Anpassungsschwierigkeiten und technische Hindernisse bemerkten. Die Intuitivität der Handhabung wurde ebenfalls positiv bewertet, trotz einiger anfänglicher Missverständnisse bezüglich der Spracherkennung.

Genauigkeit und Vollständigkeit

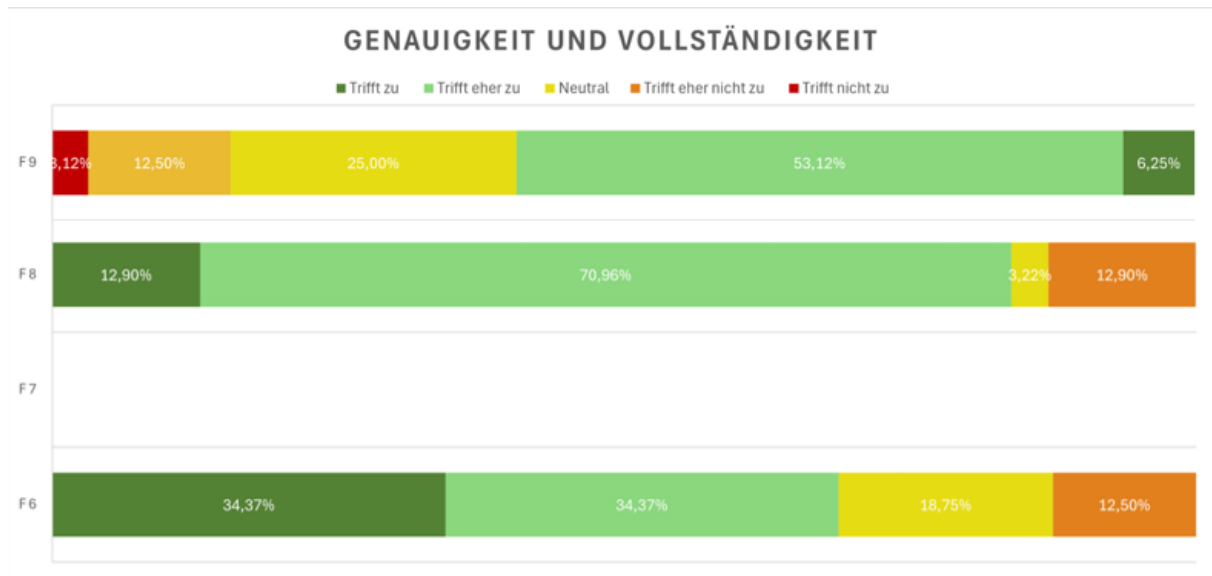


Abbildung 5.3: Auswertung der Fragen des Fragebogens mit den Themen Erleichterung, umfassendere Befunde, Genauigkeit und Korrekturen beziehungsweise Fehler. F6: Erleichtert es die Erfassung umfassender und detaillierter Befunde?, F7: Gibt es bestimmte Aspekte der Genauigkeit der Worterkennung oder der Erkennung von Abkürzungen, die Ihrer Meinung nach verbessert werden könnten?, F8: Wie zufrieden sind Sie mit der Genauigkeit der Spracherkennung und der Fähigkeit des "SpeechKit", Ihre Sprache korrekt zu transkribieren?, F9: Empfinden Sie die Diktierfehler von der Spracherkennung als betriebsverhindernd?

In Bezug auf die Unterstützung beim Erstellen umfassender und detaillierter Befunde, die in der sechsten Frage thematisiert wurde, lag die durchschnittliche Bewertung bei 2,09 (SD= 1,02). Elf Teilnehmer wählten "trifft zu", elf "trifft eher zu", sechs Personen blieben neutral und vier wählten "trifft eher nicht zu". Viele Teilnehmer fanden die Textbausteine besonders nützlich, was die Nutzung der Software erleichterte.

Trotzdem gab es auch Kritik, die in der siebten Frage, einer Freitextfrage zu Verbesserungswünschen bezüglich der Genauigkeit, deutlich wurde. Hier äußerten die Teilnehmer, dass Fachjargon manchmal falsch erkannt wurde und es bei den indischen geistlichen Schwestern aufgrund ihres Akzents zu Erkennungsproblemen kam. Zudem wurden Patientennamen nicht immer richtig erkannt, und es wurde bemängelt, dass man deutlich sprechen muss. Positiv wurde angemerkt, dass falsch erkannte Wörter nach spezifischem Training korrekt erkannt wurden. Einige Nutzer fanden es herausfordernd, im Tagesbetrieb im Pflegestützpunkt zu diktieren, da alle Anwesenden mithören konnten, was sie dazu veranlasste, die Spracherkennung ausschließlich im Nachtdienst zu nutzen.

Eine Frage die die Genauigkeit der Worterkennung thematisierte, lag die durchschnittliche Bewertung bei 2,16 (SD=0,82). Vier Personen wählten "trifft zu", 22 Teilnehmer wählten "trifft eher zu", eine Person blieb neutral und vier Teilnehmer wählten "trifft eher nicht zu". Obwohl das "Dragon Medical SpeechKit" half, Rechtschreibfehler zu vermeiden, erkannte es teilweise Wörter nicht richtig, was dazu führte, dass Nutzer Korrekturen vornehmen mussten. Dies verursachte einen ähnlichen Aufwand wie beim Tippen und der nachträglichen Rechtschreibprüfung.

Diese Problematik setzte sich auch in einer weiteren Frage fort, in der es darum ging, ob Diktierfehler als betriebsverhindernd angesehen wurden. Eine Person wählte "trifft zu", vier entschieden sich für "trifft eher zu", acht Personen blieben neutral, 17 Teilnehmer entschieden sich für "trifft eher nicht zu" und zwei für "trifft nicht zu". Die durchschnittliche Bewertung lag hier bei 3,47 (SD=0,91), wobei die Mehrheit diese Fehler nicht als gravierend einstufte.

Zusammengefasst zeigt die Auswertung, dass die Nutzung von Dragon sowohl positive als auch negative Aspekte aufweist. Die Textbausteine wurden als hilfreich bewertet, während die Erkennung von Fachjargon und Namen sowie die Handhabung in offenen Arbeitsumgebungen Herausforderungen darstellten. Trotz der gemischten Meinungen über die Genauigkeit und die damit verbundene Notwendigkeit von Korrekturen, wurde deutlich, dass Diktierfehler von der Mehrheit nicht als ernsthafte Betriebsbehinderung angesehen wurden.

Effizienz

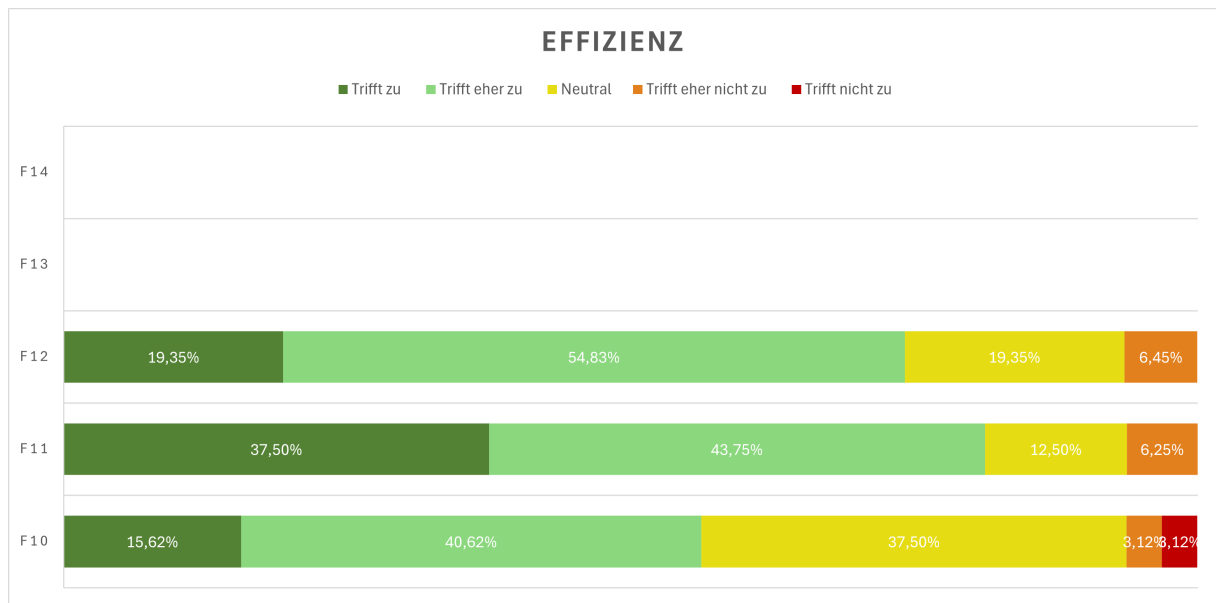


Abbildung 5.4: Auswertung der Fragen des Fragebogens mit den Themen Geschwindigkeit, Effizienzsteigerung und Fragen nach speziellen Situationen und wie sich "SpeechKit" in diesen Situationen geschlagen hat. F10: Trägt Dragon dazu bei, Fehler in der Befunderfassung zu minimieren?, F11: Hat sich die Einführung positiv auf die Dauer der Dokumentation ausgewirkt?, F12: Wie zufrieden sind Sie mit der Vielfalt der verfügbaren Textbausteine?, F13: Welchen Einfluss hat Dragon auf die Gesamteffizienz Ihrer Arbeitsabläufe?, F14: Wie beurteilen Sie die praktische Relevanz der in der Schulung vermittelten Informationen?

Hinblick auf die Minimierung von Fehlern in der Befunderfassung zeigte sich, dass eine Mehrheit der Teilnehmer entweder eine Verbesserung oder keine wesentliche Veränderung bemerkte. Die Ergebnisse deuteten darauf hin, dass die meisten entweder „trifft eher zu“ oder „neutral“ wählten, was darauf hinweist, dass die Technologie dazu beiträgt, Fehler zu reduzieren, jedoch nicht in allen Fällen signifikant.

Ein weiterer interessanter Aspekt war die Auswirkung der Spracherkennungstechnologie auf die Dauer der Dokumentation. Die Mehrheit der Teilnehmer nahm eine Verkürzung der Dokumentationszeit wahr und wählte entweder "trifft zu" oder "trifft eher zu". Dies deutet darauf hin, dass die Einführung der Technologie eine effiziente Lösung zur Zeitersparnis bei der Dokumentation darstellt.

Die Zufriedenheit mit der Vielfalt der Textbausteine zeigte, dass diese besonders für Übernahmen sehr beliebt waren. Viele Nutzer äußerten den Wunsch nach einer größeren Vielfalt solcher Bausteine, was darauf hinweist, dass die vorhandenen Textbausteine als hilfreich empfunden wurden, jedoch noch Raum für Verbesserungen besteht. Die meisten Teilnehmer gaben an, mit der Vielfalt eher zufrieden zu sein.

In Bezug auf die Gesamteffizienz der Arbeitsabläufe berichteten viele Nutzer von einer deutlichen Zeitersparnis und Effizienzsteigerung, insbesondere bei der Dokumentation von Nachsorgeprotokollen, im Nachtdienst und bei Aufnahmegesprächen. Dennoch gab es auch kritische Stimmen, die anfängliche Produktionsdämpfungen während der Einarbeitungsphase und gelegentliche Softwareausfälle aufgrund von Internetverbindungsproblemen oder Systemupdates bemängelten. Einige Nutzer waren der Meinung, dass sie mit dem Tippen schneller seien, was zeigt, dass die Technologie nicht für alle gleichermaßen effektiv war.

Abschließend wurden auch die Auswirkungen der Spracherkennungstechnologie auf die Work-Life-Balance thematisiert. Viele Teilnehmer berichteten von einer Verbesserung durch schnellere Dokumentation, die mehr Zeit für die Patienten ließ. Allerdings gaben einige an, dass die Anpassungsphase an die neue Technologie temporär zusätzlichen Arbeitsaufwand erforderte, was zu einer erhöhten Belastung führte. Dies zeigt, dass die Technologie zwar langfristig Vorteile bringt, jedoch eine anfängliche Eingewöhnungszeit erfordert.

Schulungen

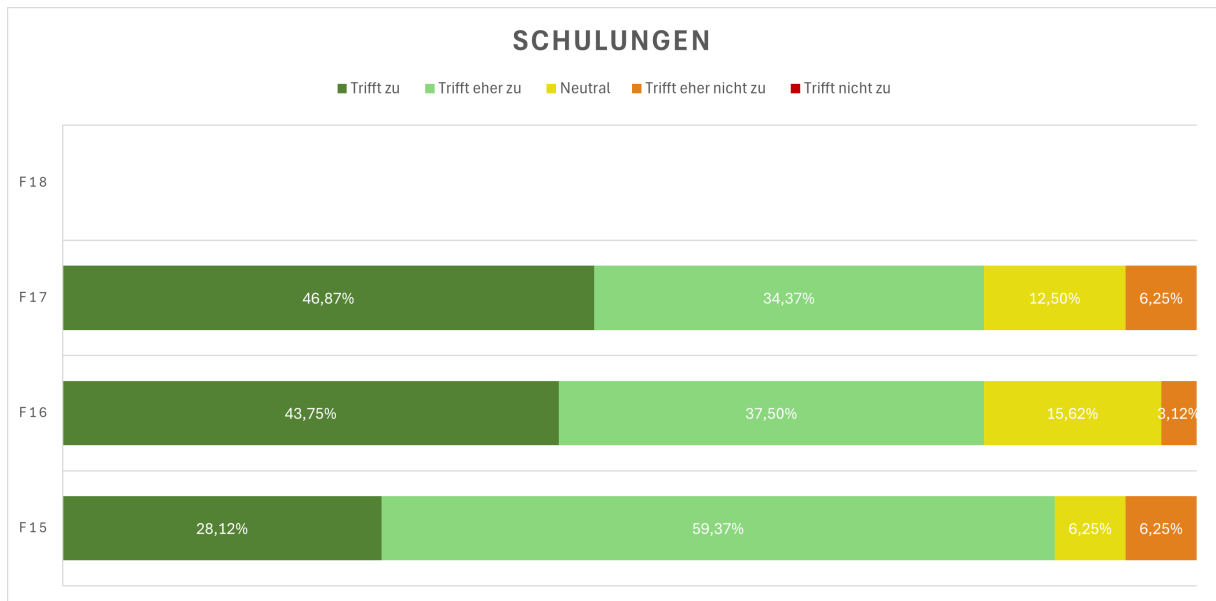


Abbildung 5.5: Auswertung der Fragen des Fragebogens mit dem Thema praktische Relevanz der Schulungen und Schulungsunterlagen. F15: Wie beurteilen Sie die praktische Relevanz der in Schulungen vermittelten Informationen?, F16: Haben Sie das Gefühl, dass die Schulungen ausreichend auf spezifische Herausforderungen in Ihrer Arbeitsumgebung eingegangen sind?, F17: Haben die Schulungen Ihre Fähigkeit zur effektiven Nutzung von Dragon verbessert?, F18: Bitte teilen Sie Ihre konkreten Gedanken oder Anmerkungen zum Umfang der Schulungsunterlagen mit und nennen Sie gegebenenfalls weitere Ressourcen, die Ihrer Meinung nach hilfreich wären bzw. Ressourcen, die nicht hilfreich waren.

Die praktische Relevanz der Schulungen wurde positiv bewertet, mit einer durchschnittlichen Bewertung von 1,91 (SD= 0,77). Neun Teilnehmer entschieden sich für "trifft zu", 19 wählten "trifft eher zu", und jeweils zwei blieben neutral oder wählten "trifft eher nicht zu". Dies zeigt, dass das vorgezeigte System mit Testpatienten leicht an einen echten Patienten anpassbar war.

Die Vorbereitung auf spezifische Herausforderungen wurde ebenfalls als hilfreich angesehen, mit einer durchschnittlichen Bewertung von 1,78 (SD= 0,83). Hier entschieden sich 14 Teilnehmer für "trifft zu" und 12 für "trifft eher zu", während fünf neutral blieben und eine Person "trifft eher nicht zu" wählte. Dabei wurden Fehlerbehandlungen hauptsächlich den Stationsleitungen mitgeteilt.

Die Verbesserung der Fähigkeit zur Nutzung der Technologie durch Schulungen erhielt ebenfalls positive Rückmeldungen. Die durchschnittliche Bewertung lag bei 1,78 (SD= 0,9), wobei 15 Teilnehmer "trifft zu" wählten, elf "trifft eher zu", vier neutral blieben

und zwei "trifft eher nicht zu" wählten. Dies deutet darauf hin, dass die Schulungen die Nutzung der Technologie effektiv unterstützten.

Die Zufriedenheit mit dem Umfang der Schulungsunterlagen zeigte jedoch Verbesserungspotenzial. Die Teilnehmer bemängelten, dass die Unterlagen nicht umfassend genug waren, und schlugen vor, zusätzliche Tutorials und spezifische Fallstudien anzubieten, um die Schulung zu verbessern. Diese Anmerkungen spiegeln das Bedürfnis nach mehr unterstützendem Material wider, um die Anwendung der Technologie noch effektiver zu gestalten.

Zufriedenheit und allgemeine Meinung

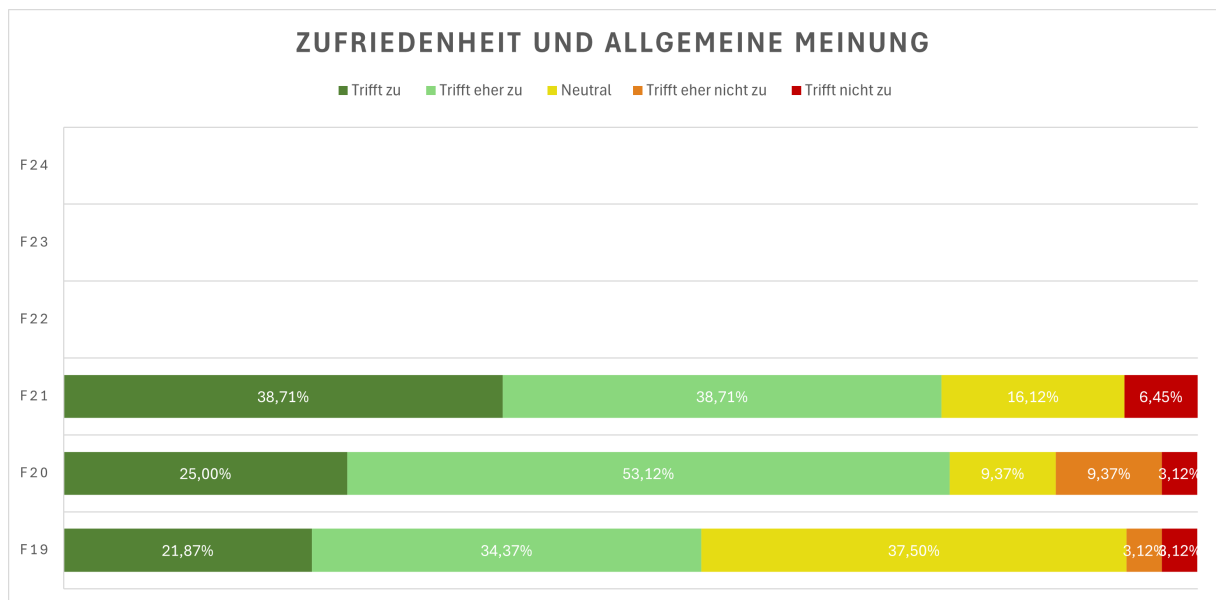


Abbildung 5.6: Auswertung der Fragen des Fragebogens mit den Themen, ob es für die Benutzer:innen mehr Vorteile als Nachteile gibt und die insgesamt Zufriedenheit über die Software. F19: Hat die Anwendung Ihre allgemeine Arbeitszufriedenheit verbessert?, F20: Gibt es für Sie mehr Vorteile als Nachteile bei der Benutzung von Spracherkennungstechnologien?, F21: Hatten Sie das Gefühl, dass das System mitlernt?, F22: Gibt es bestimmte Aspekte von Dragon, die Sie positiv oder negativ hervorheben möchten?, F23: Können Sie konkrete Beispiele für Situationen oder Herausforderungen nennen, in denen Dragon besonders hilfreich oder nicht hilfreich war?, F24: Haben Sie spezielle Anpassungen vorgenommen, um Dragon besser in Ihren Arbeitsablauf zu integrieren und wenn ja, welche?

Die Auswertung der Umfrage zeigt, dass das „Dragon Medical SpeechKit“ insgesamt mehr Vorteile als Nachteile bietet. Die durchschnittliche Bewertung lag bei 2,13 (SD= 0,96). Sieben Teilnehmer:innen stimmten voll zu, elf eher zu, zwölf blieben neutral und jeweils eine Person entschied sich für „trifft eher nicht zu“ und „trifft nicht zu“. Dies deutet

darauf hin, dass die Mehrheit der Teilnehmer:innen mehr Vorteile als Nachteile in der Nutzung der Technologie sieht.

Die Ergebnisse bezüglich der allgemeinen Arbeitszufriedenheit zeigen, dass die Technologie positiv wahrgenommen wird. Mit einer durchschnittlichen Bewertung von 2,31 (SD = 1,01) stimmten acht Teilnehmer:innen voll zu und 17 eher zu, während jeweils drei neutral blieben oder "trifft eher nicht zu" wählten und eine Person "trifft nicht zu" wählte. Die schnellere Dokumentation führte dazu, dass die meisten Teilnehmer:innen mehr Zeit für andere Aufgaben, wie die Versorgung der Patient:innen, hatten, was ihre Arbeitszufriedenheit steigerte.

Die Frage, ob das System mitlernt, wurde unterschiedlich beantwortet. Die durchschnittliche Bewertung lag bei 1,94 (SD= 1,09), was auf eine starke Abweichung hinweist. 13 Teilnehmer:innen stimmten zu, elf eher nicht zu, fünf blieben neutral und zwei wählten "trifft nicht zu". Stationen, die schon länger mit dem System arbeiteten, konnten positive Effekte bereits beobachten, während andere noch keine Verbesserungen bemerkten.

Spezifische Aspekte des „Dragon Medical SpeechKit“ wurden sowohl positiv als auch negativ hervorgehoben. Die Genauigkeit und die Textbausteine der Spracherkennung wurden als besonders positiv bewertet, was zu einer verbesserten Effizienz in der Dokumentation führte. Andere schätzten die einfache Integration in den Pflegebericht, sodass keine zusätzlichen Applikationen geöffnet werden müssen. Negativ bemerkt wurde, dass teilweise Nutzernamen geändert werden mussten, anfängliche Lizenzfehler auftraten und nach einem Systemupdate die Spracherkennungstechnologie einfrierte. Zudem wurde kritisiert, dass die Umgebung beim Diktieren mithören kann.

Konkrete Beispiele für Herausforderungen oder hilfreiche Situationen verdeutlichen, dass die Spracherkennung auch komplexe Dokumentationen problemlos bewältigte. Die Textbausteine ermöglichten schnellere Übernahmen und waren in Notfällen hilfreich. Herausforderungen traten vor allem in lauter Umgebung, bei zu weit entfernten Mikrofonen oder bei der Interaktion mit personenspezifischem Vokabular auf. Es brauchte Zeit, bis das System die Wörter registriert hatte.

Zuletzt wurde untersucht, ob Benutzer:innen spezielle Anpassungen vorgenommen haben, um „Dragon Medical SpeechKit“ besser in ihren Arbeitsablauf zu integrieren. Einige teilten mit, dass sie den Aufnahmeprozess umstrukturiert hatten, die Mehrheit hatte jedoch noch keine Anpassungen umgesetzt. Diese Ergebnisse zeigen, dass die Technologie sowohl positive als auch negative Aspekte aufweist, und dass die Nutzererfahrungen je nach Anwendung und Umgebung variieren.

Insgesamt wurde die Integration von "Dragon Medical Speechkit", überwiegend positiv bewertet. Die Benutzerfreundlichkeit wurde von 70,63 % der Befragten als trifft zu und

von 10,49 % als trifft eher zu eingestuft. Dies zeigt, dass die Mehrheit der Nutzer die Technologie als intuitiv und leicht zu bedienen empfindet. Nur 8,7 % der Befragten äußerten sich negativ zur Benutzerfreundlichkeit. In Bezug auf Genauigkeit und Vollständigkeit gaben 46,60 % der Befragten an, dass diese Kriterien voll zutreffen, während 27,84 % eher zustimmten. Dies bedeutet, dass eine Mehrheit von über 74 % die Spracherkennung als präzise und umfassend wahrnimmt. Ein kleinerer Anteil von 8,73 % der Nutzer war der Meinung, dass die Technologie in diesen Bereichen nicht vollständig überzeugt, was auf Herausforderungen mit Erkennungsfehler hindeutet. Die Effizienz der Befunderfassung wurde von 28,57 % der Befragten als sehr gut und von 49,21 % als eher gut bewertet. Dies zeigt, dass fast 78 % der Nutzer eine Verbesserung der Effizienz durch die Nutzung von "Dragon Medical Speechkit" feststellten. Besonders die Möglichkeit der Verfassung von umfassenderen Berichten wurde positiv betont. Ein kleinerer Anteil von 6,35 % empfand die Effizienz jedoch als unzureichend, was darauf hindeutet, dass es noch weiteren Verbesserungspotenzial gibt. Bei der Bewertung der Schulungen gaben 33,93 % der Befragten an, dass diese sehr gut sind und 37,50 % stimmten eher zu. Dies deutet darauf hin, dass die Schulungen insgesamt positiv wahrgenommen werden, obwohl fast zehn Prozent der Befragten Verbesserungsbedarf sahen. Hier hätte eine intensivere oder besser strukturierte Schulung zur weiteren Verbesserung beitragen können. Die allgemeine Zufriedenheit bei der Integration von "Dragon Medical Speechkit" wurde von 28,28 % der Befragten als sehr gut und von 39,36 % als eher gut bewertet. Damit sind über 67 % der Nutzer zufrieden mit der Technologie. Dennoch gaben 12,12 % der Befragten an, dass sie nicht zufrieden sind, was auf bestehende Herausforderungen bei der Integration oder Usernamen wechsel der Benutzer/innen hinweist.

Zusammenfassend zeigt die Umfrage, dass die Integration von "Dragon Medical Speech-Kit" im Gesundheitswesen überwiegend positiv aufgenommen wird. Besonders hervorgehoben werden die Benutzerfreundlichkeit und Effizienz der Software. Auch die Genauigkeit und Vollständigkeit werden positiv bewertet, wenngleich hier noch Optimierungspotential besteht. Ein weiterer Bereich, der Verbesserungen benötigt, sind die Schulungen. Insgesamt äußern sich die meisten Nutzer jedoch zufrieden mit "Dragon Medical SpeechKit".

Kapitel 6

Diskussion

6.1 Zusammenfassung

Diese Bachelorarbeit untersucht die Einführung und Nutzung der Spracherkennungstechnologie "Dragon Medical SpeechKit" im klinischen Umfeld des Krankenhauses der Barmherzigen Brüder in Graz. Im Mittelpunkt stehen die Auswirkungen dieser Technologie auf die Qualität, Effizienz und Vollständigkeit der medizinischen Dokumentation. Durch umfangreiche Literaturrecherche und eine empirische Befragung des medizinischen Personals wird die Akzeptanz, sowie die praktische Anwendung der Technologie bewertet. Diese Technologie entlastet das medizinische Personal und schafft laut den Nutzer:innen mehr Zeit für die Patientenversorgung. Die Mehrheit der Nutzer (über 74 %) nahm die Spracherkennung als präzise und umfassend wahr, und fast 78 % der Nutzer berichteten von einer Verbesserung der Effizienz bei der Nutzung des „Dragon Medical SpeechKit“. Besonders positiv wurde die Möglichkeit hervorgehoben, umfassendere Berichte zu verfassen. Die Schulungen zur Nutzung der Technologie wurden insgesamt positiv bewertet. Allerdings sahen fast zehn Prozent der Befragten Verbesserungsbedarf, was darauf hindeutet, dass intensivere oder besser strukturierte Schulungen zur weiteren Verbesserung beitragen könnten. Mehrere technische Schwierigkeiten wurden identifiziert, darunter Probleme mit der Lizenzierung, technische Anforderungen der Hardware und die Anpassung an die Netzwerkinfrastruktur. Diese Herausforderungen unterstreichen die Bedeutung sorgfältiger Planung und Anpassung der Technologie an die spezifischen Bedingungen und Bedürfnisse des klinischen Umfelds. Die Mehrheit der Nutzer nahm die Technologie positiv auf, wobei die intuitive Bedienung und die Erleichterung im täglichen Arbeitsablauf hervorgehoben wurden. Jedoch gaben 12,12 % der Befragten an, dass sie nicht zufrieden sind, was auf bestehende Herausforderungen bei der Integration oder auf spezifische Benutzerprobleme hinweist. Abschließend wird die Notwendigkeit kontinuierlicher Forschung und Entwicklung reflektiert, um die Integration und Effektivität von Spracher-

kennungstechnologien im Gesundheitswesen zu optimieren. Es wird betont, dass trotz der offensichtlichen Vorteile eine laufende Anpassung und Bewertung erforderlich sind, um die Technologie effektiv in die klinische Praxis zu integrieren und ihre Akzeptanz unter den medizinischen Fachkräften zu sichern. In dieser Arbeit wurden drei Untersuchungsmethoden angewendet: Literaturrecherche, Durchführung einer Beobachtungsanalyse, eine Umfrage zur Nutzerzufriedenheit und ein Interview. Es wurden zwei Forschungsfragen beantwortet. Die erste Frage, die sich auf die Auswirkungen der Integration von "Dragon" auf die Qualität der Befunderfassung und die damit verbundenen Herausforderungen konzentrierte, wurde teils während der Beobachtungsanalyse durch gezielte Beobachtung der Dokumentationsgeschwindigkeit und -vollständigkeit und teils durch den Fragebogen aus Nutzersicht beantwortet. Ein Mitarbeiter des IT-Business Management gab Auskunft über technische Herausforderungen während der Implementierung. Die zweite Forschungsfrage bezog sich auf die Auswirkungen der Implementierung vom "Dragon Medical SpeechKit" auf die Akzeptanz der Technologie, insbesondere hinsichtlich Nutzerfreundlichkeit und Anwenderzufriedenheit. Die Beobachtungsanalyse und der Fragebogen zeigten, dass "Dragon Medical SpeechKit" positiv aufgenommen wurde, wobei die Mehrheit der Befragten die intuitive Bedienung und die Erleichterung im täglichen Arbeitsablauf hervorhob.

6.2 Limitierung

Bei der Durchführung der Beobachtungsanalyse und der Umfragen traten mehrere Einschränkungen auf. Die Studie war zeitlich begrenzt und fand nur an ausgewählten Stationen statt, die teilweise bereits Erfahrungen mit dem "Dragon Medical SpeechKit" hatten. Dabei ist anzunehmen, dass bei neu beginnenden Stationen die anfänglichen Akzeptanzprobleme größer sind. Zudem ergaben sich Schwierigkeiten bei der Literaturrecherche zu den Nachteilen des "Dragon Medical SpeechKit" sowie von Spracherkennungssystemen allgemein. Obwohl kritische Meinungen und Eindrücke zur Spracherkennung gefunden wurden, war die spezifische Kritik am "Dragon Medical SpeechKit" durchgehend positiv. Ein weiteres Problem stellte die Qualität der Spracherkennung bei starken Umgebungsgläuschen dar. Außerdem war die allgemeine Akzeptanz des Systems gering, da viele Nutzer die vertrauten Systeme bevorzugten und sich nicht auf das neue Spracherkennungssystem einlassen wollten. Die Antworten der Befragten des Fragebogens könnten durch subjektive Wahrnehmungen verzerrt sein, da sie auf Selbsteinschätzungen beruhten. Es besteht auch die Gefahr von Antworttendenzen, wie die Tendenz zur Mitte oder sozial erwünschte Antworten, was die Genauigkeit der Ergebnisse beeinträchtigen kann. Die Größe wie auch die Altersverteilung der Stichprobe ist möglicherweise nicht repräsentativ

für alle Nutzer:innen, was die Generalisierbarkeit der Ergebnisse einschränkt. Weiterhin könnten die Beobachtungen subjektiv und von den Interpretationen der Autorin beeinflusst sein, was zu einer Verzerrung der Ergebnisse führen kann. Die Beobachtungsanalyse wurde nur über einen begrenzten Zeitraum durchgeführt, was möglicherweise nicht alle Aspekte der Technologienutzung erfasst.

6.3 Reflexion und Zukunftsausblick

Die Forschung hat gezeigt, dass die Einführung von Spracherkennungstechnologien wie "Dragon Medical SpeechKit" zu einer effizienteren und umfassenderen medizinischen Dokumentation geführt hat. Die Untersuchung offenbarte jedoch auch bedeutende Einschränkungen. Der Datenschutz personenbezogener Daten bleibt ein kritisches Thema. Dragon ist eine Cloud-basierte Lösung, was bedeutet, dass die Daten extern gespeichert werden. Die Spracherkennung wird von Microsoft Azure gehostet, einer Cloud-Plattform mit HITRUST CSF-Zertifizierung, die strenge Sicherheitsstandards erfüllt. Die Lösung ist DSGVO-, BDSG- und C5-konform (Cloud Computing Compliance Criteria Catalogue), der vom Bundesamt für Sicherheit in der Informationstechnik (BSI) in Deutschland entwickelt wurde und spezifische Anforderungen und Empfehlungen für die Sicherheit von Cloud-Diensten definiert. Zudem ist die Lösung nach ISO 27001 zertifiziert. Zudem ist es notwendig, dem medizinischen Personal umfangreiche Schulungen anzubieten und die Benutzer über die Vorteile der Technologie aufzuklären. Es ist ebenfalls wichtig, regelmäßig Feedback von den Nutzer:innen zu sammeln und, wenn möglich, Vorschläge in die Weiterentwicklung der Technologie zu integrieren. Die Zusammenarbeit mit Techniker:innen, Ärzt:innen und Pflegepersonal hat die Bedeutung einer multidisziplinären Perspektive auf die Spracherkennung unterstrichen. Sie hat gezeigt, wie entscheidend regelmäßige Meetings und Kommunikation sind, um Missverständnisse zu vermeiden. Es wird eine kontinuierliche Verbesserung durch den fortschreitenden Einsatz künstlicher Intelligenz erwartet, auch wenn die Innovationsrate möglicherweise abnimmt. Maschinelles Lernen, insbesondere semi-supervised und self-supervised Learning, wird zunehmend wichtig, um effektiver aus beschrifteten und unbeschrifteten Daten zu lernen. Die Personalisierung und die Fähigkeit der Systeme, sich an sprachliche Merkmale der Nutzer anzupassen, werden verbessert. Fortschritte in der Emotionserkennung könnten die Patient:innen Interaktionen und das Gesundheitswesen verbessern. Zudem gewinnen ethische Überlegungen und Datenschutz an Bedeutung, da die Systeme immer mehr persönliche Daten erfassen. Bei der Einführung weiterer technischer Programme ist es wichtig, die Benutzerakzeptanz gezielt zu fördern und die Technologie weiter an die spezifischen Bedürfnisse des me-

dizinischen Personals anzupassen. Zukünftig könnte der Einsatz drahtloser Mikrofone eine weitere Erleichterung darstellen. Die Fortführung der Forschung ist in diesem Bereich essenziell, insbesondere in Bezug auf die Herausforderung des Datenschutzes und der Datensicherheit. Datenschutz und Datensicherheit sind bei der Spracherkennung besonders wichtig, da sensible personenbezogene Daten verarbeitet werden. Die Lösung ist DSGVO-konform, was bedeutet, dass höchste Datenschutz- und Sicherheitsanforderungen garantiert sind. Die Daten werden in Microsoft Azure, einem sicheren Rechenzentrum, gespeichert, das den gesetzlichen Anforderungen und den spezifischen Anforderungen des Gesundheitssystems entspricht. Weiterhin wird es wichtig sein, die Systeme so zu gestalten, dass sie eine noch breitere Palette von Dialekten und Akzenten zuverlässig erkennen können. Zusätzlich können diese Systeme laut [Badawy et al., 2023] durch KI profitieren, um aus Patientenakten zu lernen, Muster zu erkennen und prädiktive Analysen zu bieten. Diese können helfen, potenzielle gesundheitliche Probleme frühzeitig zu erkennen und präventive Maßnahmen einzuleiten. Weiters könnten durch maschinelles Lernen die Spracherkennungstechnologien besser auf individuelle Sprechmuster und Fachterminologien der Benutzer abgestimmt werden, was die Genauigkeit weiter erhöhen und auch die Akzeptanz und Zufriedenheit der Anwender:innen verbessern würde.

Zusammenfassend bietet die Spracherkennung im Gesundheitswesen erhebliche Vorteile, insbesondere in Bezug auf Effizienz und Dokumentationsqualität. Die kontinuierliche Verbesserung und Anpassung der Technologie an die Bedürfnisse der Nutzer:innen sind entscheidend, um ihre Akzeptanz und Wirksamkeit zu maximieren.

Literaturverzeichnis

- 3M (2024). MModal medizinische Spracherkennung. URL: https://www.3m.com/3M/en_US/health-information-systems-us/create-time-to-care/clinician-solutions/speech-recognition/fluency-direct/. Abgerufen am 25.03.2024.
- Alanazi, H., Abdullah, A., und Qureshi, K. (2017). A critical review for developing accurate and dynamic predictive models using machine learning methods in medicine and health care. *Journal of medical systems*, 41:1–10. DOI: 10.1007/s10916-017-0715-6.
- Aldosari, B., Babsai, R., Alanazi, A., Aldosari, H., und Alanzi, A. (2023). The progress of speech recognition in health care: Surgery as an example. *Healthcare Transformation with Informatics and Artificial Intelligence*, 305:414. DOI: 10.3233/SHTI230519.
- AWS (2024). Amazon transcribe medical. URL: <https://aws.amazon.com/de/transcribe/medical/>. Abgerufen am 25.03.2024.
- Badawy, M., Ramadan, N., und Hefny, H. (2023). Healthcare predictive analytics using machine learning and deep learning techniques: a survey. *Journal of Electrical Systems and Information Technology*, 10(1):40. DOI: 10.1186/s43067-023-00108-y.
- Bahl, L., Brown, P., de Souza, P., und Picheny, M. (1988). Acoustic markov models used in the tangora speech recognition system. In *ICASSP-88., International Conference on Acoustics, Speech, and Signal Processing*, pages 497–498. IEEE Computer Society. DOI: 10.1109/ICASSP.1988.196628.
- Bajpai, M. und Sharda, P. (2022). Speech recognition technology for opac service: An innovative idea for indian libraries. URL: https://www.researchgate.net/publication/351991901_SPEECH_RECOGNITION_TECHNOLOGY_FOR_OPAC_SERVICE_AN_INNOVATIVE_IDEA_FOR_INDIAN_LIBRARIES. Aufgerufen am 18.03.2024.
- Bakst, S., Yilmaz, E., und Castán, D. (2023). Towards improving automatic speech recognition for underrepresented dialects with data augmentation. *The Journal of the Acoustical Society of America*. DOI: 10.1121/10.0018920.

- Benkerzaz, S., Elmir, Y., und Dennai, A. (2019). A study on automatic speech recognition. *Journal of Information Technology Review*, 10(3):77–85. DOI: 10.6025/jitr/2019/10/3/77-85.
- Blackley, S., Huynh, J., Wang, L., Korach, Z., und Zhou, L. (2019). Speech recognition for clinical documentation from 1990 to 2018: a systematic review. *Journal of the american medical informatics association*, 26(4):324–338. DOI: 10.1093/jamia/ocy179.
- Blackley, S., Schubert, V., Goss, F., Assad, W., Garabedian, P., und Zhou, L. (2020). Physician use of speech recognition versus typing in clinical documentation: A controlled observational study. *International Journal of Medical Informatics*, 141:104178. DOI: 10.1016/j.ijmedinf.2020.104178.
- Brandon, J. (2021). Using unethical data to build a more ethical world: How callminer handles imperfections in speech recognition. *AI and Ethics*, 1(2):101–108. DOI: 10.1007/s43681-020-00006-3.
- Byrne, A. (2023). Case study: automated transcription of hospital letters with 3m m* modal speech-recognition software. *British Journal of Nursing*, 32(Sup16b):S10–S15. DOI: 10.12968/bjon.2023.32.Sup16b.S10.
- Carrón, J., Campos-Roca, Y., Madruga, M., und Perez, C. (2021). A mobile-assisted voice condition analysis system for parkinson’s disease: Assessment of usability conditions. *Biomedical engineering online*, 20:1–24. DOI: 10.1186/s12938-021-00951-y.
- Carstensen, K.-U., Ebert, C., Ebert, C., Jekat, S., Klabunde, R., und Langer, H. (2010). *Computerlinguistik und Sprachtechnologie: Eine Einführung*. Spektrum Akademischer Verlag Heidelberg, 3 edition. ISBN=978-3-8274-2023-7.
- Chassagnon, G., Vakalopoulou, M., Paragios, N., und Revel, M.-P. (2020). Deep learning: definition and perspectives for thoracic imaging. *European radiology*, 30:2021–2030. DOI: 10.1007/s00330-019-06564-3.
- Chen, M. und Decary, M. (2020). Artificial intelligence in healthcare: An essential guide for health leaders. 33:10–18. DOI: 10.1177/0840470419873123.
- Cieri, C., Miller, D., und Walker, K. (2004). The fisher corpus: A resource for the next generations of speech-to-text. URL: <http://www.lrec-conf.org/proceedings/lrec2004/pdf/767.pdf>. Augerufen am 18.03.2024.

- Darcy, A., Beaudette, A., Chiauzzi, E., Daniels, J., Goodwin, K., Mariano, T., Wicks, P., und Robinson, A. (2022). Anatomy of a woebot(wb001): agent guided cbt for women with postpartum depression. 19(4):287–301. DOI: 10.1080/17434440.2022.2075726.
- Darnell, T. (1997). Devices as dynamic as the people who use them,. URL: <https://elibrary.ru/item.asp?id=2948041>. Abgerufen am 18.03.2024.
- datasolut (2024). Machine learning vs. deep learning: Wo ist der unterschied? URL: <https://datasolut.com/machine-learning-vs-deep-learning/>. Abgerufen am 09.04.2024.
- Deepa, P. und Khilar, R. (2022). Speech technology in healthcare. *Measurement: Sensors*, 24:100565. DOI: 10.1016/j.measen.2022.100565.
- Dettmeyer, R. (2006). *Medizinische Maßnahmen zum Zwecke der Beweissicherung*. Springer Berlin Heidelberg. ISBN=978-3-540-29864-9.
- Do, H. M., Sheng, W., Harrington, E., und Bishop, A. (2020). Clinical screening interview using a social robot for geriatric care. *IEEE Transactions on Automation Science and Engineering*, 18(3):1229–1242. DOI: 10.1109/TASE.2020.2999203.
- Eggert, M. und Kreuzer, V. (2022). Sprachsteuerung im gesundheitswesen–anforderungen und auswahl geeigneter anbieter. *HMD Praxis der Wirtschaftsinformatik*, 59(6):1640–1652. DOI: 10.1365/s40702-022-00919-z.
- EU-Parlament (2016). Rechtsnormen der dsgvo. URL: <https://www.datenschutz-grundverordnung.eu/>. Abgerufen am 14.03.2024.
- Everett, M., Redner, J., Kalenscher, A., Durso, D., und Nguyen, S. (2023). Speech recognition technology for increasing nursing documentation efficiency. URL: <https://www.himss.org/resources/speech-recognition-technology-increasing-nursing-documentation-efficiency>. Abgerufen am 09.02.2024.
- Fishman, E. (1997). Speech-recognition software improves, gains recognition. URL: <https://go.gale.com/ps/i.do?id=GALE%7CA19213568&sid=googleScholar&v=2.1&it=r&linkaccess=abs&iissn=00011843&p=HRCA&sw=w&userGroupName=anon%7E99914459&aty=open-web-entry>. Abgerufen am 21.03.2024.
- Gaikwad, S., Bharti, G., und Yannawa, P. (2010). A review on speech recognition technique. *International Journal of Computer Applications*, 10:1. DOI: 10.5120/1462-1976.

- Goss, F., Blackley, S., Ortega, C., Kowalski, L., Landman, A., Lin, C.-T., Meter, M., Bakes, S., Gradwohl, S., Bates, D., und Zhou, L. (2019). A clinician survey of using speech recognition for clinical documentation in the electronic health record. *International journal of medical informatics*, 130:103938. DOI: 10.1016/j.ijmedinf.2019.07.017.
- Gressner, A. und Gressner, O. (2019). *Befunderstellung, Teilschritte*, pages 412–413. Springer Berlin Heidelberg, Berlin, Heidelberg. ISBN=978-3-662-48986-4.
- Haas, P. (2005). Grundlegende aspekte der medizinischen dokumentation und organisation. *Medizinische Informationssysteme und Elektronische Krankenakten*, pages 113–184. DOI: 10.1007/3-540-26855-3₃.
- Hoffman, T. (1997). Techno-phobic mds refuse to say ‘ah!’(doctors’ reluctance to use computer technology for making patient records). URL: https://archive.org/details/sim_computerworld_1997-05-12_31_19/page/n75/mode/2up. Augerufen am 21.03.2024.
- Hopkins, N. (2024). Speech recognition technology in healthcare. URL: <https://www.probecx.com/en-au/blog/speech-recognition-technology-in-healthcare>.
- Hoy, M. (2018). Alexa, siri, cortana, and more: An introduction to voice assistants. *Medical Reference Services Quarterly*, 37(1):81–88. DOI: 10.1080/02763869.2018.1404391.
- Jia, P. und Yang, H. (2022). Speech assistant system for anesthesia surgery based on intelligent decision algorithm. In *2022 2nd International Conference on Bioinformatics and Intelligent Computing*, pages 345–350. DOI: 10.1145/3523286.3524567.
- Johnson, M., Lapkin, S., Long, V., Sanchez, P., Suominen, H., Basilakis, J., und Dawson, L. (2014). A systematic review of speech recognition technology in health care. *BMC medical informatics and decision making*, 14:1–14. DOI: 10.1186/1472-6947-14-94.
- Jongman, S., Khoe, Y., und Hintz, F. (2021). Vocabulary size influences spontaneous speech in native language users: Validating the use of automatic speech recognition in individual differences research. *Language and Speech*, 64(1):35–51. DOI: 10.1177/0023830920911079.
- Joshi, S., Kumari, A., Pai, P., und Sangaonkar, S. (2017). Voice recognition system. *Journal for Research*, 3(01). DOI: 10.1109/J4RV3I1005.
- Jurafsky, D. und Martin, J. (2024). Speech and language processing. URL: <https://web.stanford.edu/~jurafsky/slp3/ed3book.pdf>. Augerufen am 18.03.2024.

- Jyothi, P., Johnson, L., Chelba, C., und Strope, B. (2012). Distributed discriminative language models for google voice-search. In *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5017–5020. IEEE. DOI: 10.1109/ICASSP.2012.6289047.
- Kim, D. und Leeming, B. (1990). Computerized on-line reporting of coronary angiograms with the clip system. *AJR. American journal of roentgenology*, 154(6):1307–1310. DOI: 10.2214/ajr.154.6.2110746.
- Kim, D., Oh, J., Im, H., Yoon, M., Park, J., und Lee, J. (2021a). Automatic classification of the korean triage acuity scale in simulated emergency rooms using speech recognition and natural language processing: a proof of concept study. *Journal of Korean Medical Science*, 36(27). DOI: 10.3346/jkms.2021.36.e175.
- Kim, S., Mir-Bashiri, S., Matthies, P., Sommer, W., und Nörenberg, D. (2021b). Integration der strukturierten befundung in den radiologischen routine-workflow. *Der Radiologe*, 61(11):1005. DOI: 10.1007/s00117-021-00917-0.
- Kuczera, L. (2021). Qualität beim schriftdolmetschen mittels spracherkennung: Modelle zur qualitätsbewertung für das interlinguale respeaking und das simultandolmetschen im vergleich: a comparison of models for quality assessment in interlingual respeaking and simultaneous interpreting. URL: <https://unipub.uni-graz.at/obvugrhs/content/titleinfo/5749722/full.pdf>. Augerufen am 18.03.2024.
- Kudryavtsev, N., Bardasova, K., und Khoruzhaya, A. (2023). Speech recognition technology in radiology. *Digital Diagnostics*, 4(2):185–196. DOI: 10.17816/DD321420.
- Kumar, Y. (2024). A comprehensive analysis of speech recognition systems in healthcare: Current research challenges and future prospects. *SN Computer Science*, 5(1):137. DOI: 10.1007/s42979-023-02466-w.
- Latif, S., Qadir, J., Qayyum, A., Usama, M., und Younis, S. (2020). Speech technology for healthcare: Opportunities, challenges, and state of the art. *IEEE Reviews in Biomedical Engineering*, 14:342–356. DOI: 10.1109/RBME.2020.3006860.
- Le Prell, C. und Clavier, O. (2017). Effects of noise on speech recognition: Challenges for communication by service members. *Hearing research*, 349:76–89. DOI: 10.1016/j.heares.2016.10.004.

- Lee, W., Seong, J., Ozlu, B., Shim, B., Marakhimov, A., und Lee, S. (2021). Biosignal sensors and deep learning-based speech recognition: A review. *Sensors*, 21(4):1399. DOI: 10.3390/s21041399.
- Lenarz, T. und Boenninghaus, H.-G. (2012). *Hals-Nasen-Ohren-Heilkunde*. Springer-Verlag. ISBN=978-3-642-21130-0.
- Li, B., Sainath, T., Narayanan, A., Caroselli, J., Bacchiani, M., Misra, A., Shafran, I., Sak, H., Pundak, G., Chin, K., Sim, K., Weiss, R., Wilson, K., Variani, E., Kim, C., Siohan, O., Weintraub, M., McDermott, E., Rose, R., und Shannon, M. (2017). Acoustic modeling for google home. In *Interspeech*, pages 399–403. DOI: 10.21437/Interspeech.2017-234.
- Lobmeier, S. (2020). Kontron entwickelt sprachsteuerung für op-roboterarm. URL: <https://www.elektroniknet.de/medizintechnik/systeme-anwendungen/kontron-entwickelt-sprachsteuerung-fuer-op-roboterarm.173489.html>. Abgerufen am 09.02.2024.
- Lowerre, B. (1976). The harpy speech recognition system. URL: <https://stacks.stanford.edu/file/druid:rq916rn6924/rq916rn6924.pdf>. Abgerufen am 09.04.2024.
- Magrabi, F., Hodgson, T., und Coiera, E. (2018). Evaluating the usability of speech recognition to create clinical documentation using a commercial electronic health record. *International Journal of Medical Informatics*, 113:38–42. DOI: 10.1016/j.ijmedinf.2018.02.011.
- Malik, M., Malik, M., Mehmood, K., und Makhdoom, I. (2021). Automatic speech recognition: a survey. *Multimedia Tools and Applications*, 80:9411–9457. DOI: 10.1007/s11042-020-10073-7.
- Mandel, M. (1992). A commercial large-vocabulary discrete speech recognition system: Dragondictate. *Language and speech*, 35(1-2):237–246. DOI: 10.1177/002383099203500218.
- MIMIC (2021). Mimic documentation. URL: <https://mimic.mit.edu/docs/>. Abgerufen am 14.03.2024.
- Moore, M., Papreja, P., Saxon, M., Berisha, V., und Panchanathan, S. (2020). Uncommonvoice: A crowdsourced dataset of dysphonic speech. URL:

- https://img1.wsimg.com/blobby/go/bb8819fe-ceab-4aab-9326-de58f46295cf/downloads/UncommonVoice_IS2020.pdf?ver=1702936360906. Abgerufen am 14.03.2024.
- Mozilla (2024). Datensätze. URL: <https://commonvoice.mozilla.org/de/datasets>. Abgerufen am 14.03.2024.
- Muhammad, G. (2015). Automatic speech recognition using interlaced derivative pattern for cloud based healthcare system. *Cluster Computing*, 18:795–802. DOI: 10.5120/1462-1976.
- Nuance (2010). Nuance dragon naturally speaking 11. https://www.nuance.com/content/dam/nuance/en_us/collateral/corporate/company-policies/accessibility/cp-section-508-vpat-nuance-dns-v11-detailed-r1.doc-en-us.pdf. Abgerufen am 09.04.2024.
- Nuance (2016). Spracherkennung für jede medizinische anwendung auf jedem gerät, an jedem ort, zu jeder zeit. URL: https://www.nuance.com/content/dam/nuance/de_de/collateral/healthcare/data-sheet/ds-dragon-medical-speechkit-de-de.pdf. Abgerufen am 09.04.2024.
- Nuance (2024). Dragon. URL: <https://www.nuance.com/de-de/healthcare/provider-solutions/speech-recognition/dragon-medical-one.html>. Abgerufen am 09.04.2024.
- Olive, J., Christianson, C., und McCary, J. (2011). *Handbook of natural language processing and machine translation: DARPA global autonomous language exploitation*. Springer Science & Business Media. ISBN=978-1-4419-7712-0.
- OM Digital Solutions (2024). Odms. URL: <https://dictation.omsystem.com/de/>. Abgerufen am 25.03.2024.
- Ondáš, S., Kiktova, E., Pleva, M., Oravcova, M., Hudack, L., Juhar, J., und Zimmermann, J. (2020). Pediatric speech audiometry web application for hearing detection in the home environment. *Electronics*, 9(6):994. DOI: 10.3390/electronics9060994.
- Onitilo, A., Shour, A., Puthoff, D., Tanimu, Y., Joseph, A., und Sheehan, M. (2023). Evaluating the adoption of voice recognition technology for real-time dictation in a rural healthcare system: A retrospective analysis of dragon medical one. *Plos one*, 18(3):e0272545. DOI: 10.1371/journal.pone.0272545.

- Pallett, D. (1989). Benchmark tests for darpa resource management database performance evaluations. pages 536–539. DOI: 10.1109/ICASSP.1989.266482.
- Pereira, A., Folgado, D., Cotrim, R., und Sousa, I. (2019). Physiotherapy exercises evaluation using a combined approach based on semg and wearable inertial sensors. In *Biosignals*, pages 73–82. DOI: 10.5220/0007391300730082.
- Philips (2024). Speech mike. URL: <https://www.bechtle.com/at/shop/philips-speechmike-premium-touch-3720--4242080--p>. Abgerufen am 09.04.2024.
- Pieraccini, R. (2012). From audrey to siri. *Is speech recognition a solved problem*, 23. DOI: 10.1109/11288.
- Pompili, A., Solera-Ulrena, R., Abad, A., Cardoso, R., Guimaraes, I., Fabbri, M., Martins, I., und Ferreira, J. (2020). Assessment of parkinson’s disease medication state through automatic speech analysis. *arXiv preprint arXiv:2005.14647*. DOI: 10.48550/arXiv.2005.14647.
- Raphael, L., Borden, G., und Harris, K. (2021). *Speech Science Primer: Physiology, Acoustics, and Perception of Speech*. Lippincott Williams, Wilkins, 5 edition. ISBN=92DD-K21-SPJR.
- Rebala, G., Ravi, A., und Churiwala, S. (2019). *An Introduction to Machine Learning*. Springer. ISBN=978-3-030-15729-6.
- Rippel, J. (2021). *Spracherkennung für Autoren*. epubli. ISBN=9783753156583.
- RIS (2023). Bundesrecht konsolidiert: Gesamte rechtsvorschrift für datenschutzgesetz, fassung vom 14.03.2024. URL: <https://www.ris.bka.gv.at/GeltendeFassung.wxe?Abfrage=Bundesnormen&Gesetzesnummer=10001597>. Abgerufen am 14.03.2024.
- Ronaldo, P., Kock, N., und Sonsini, J. (2004). An analysis of the implementation and impact of speech-recognition technology in the healthcare sector. *Perspectives in Health Information Management/AHIMA, American Health Information Management Association*, 1. DOI: 10.1007/PMC2047322.
- Saxena, K., Diamond, R., Conant, R., Mitchell, T., Gallopyn, G., und Yakimow, K. (2018). Provider adoption of speech recognition and its impact on satisfaction, documentation quality, efficiency, and cost in an inpatient ehr. *AMIA Summits on Translational Science Proceedings*, 2018:186. DOI: 10.1007/PMC5961784.

- Schnell, R., Hill, P., und Esser, E. (1999). *Methoden der empirischen Sozialforschung*, volume 6. R. Oldenbourg München. ISBN=978-3-531-17809-7.
- Schukat-Talamazzini, E. G. (2013). *Automatische Spracherkennung: Grundlagen, statistische Modelle und effiziente Algorithmen*. Springer-Verlag. ISBN 978-3-322-96180-8.
- Sokolov, S., Alimov, O., Tyapkin, D., Katorin, Y., und Mosieev, A. (2020). Modern social engineering voice cloning technologies. In *2020 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus)*, pages 513–516. DOI: 10.1109/EIConRus49466.2020.9038954.
- Song, Z. (2020). English speech recognition based on deep learning with multiple features. *Computing*, 102(3):663–682. DOI: 10.1007/s00607-019-00753-0.
- Sonix (2024). A short history of speech recognition. URL: <https://sonix.ai/history-of-speech-recognition>. zuletzt besucht am 27.1.2024.
- Sosnowsky, R. (2023). Deshalb ist die bedienung mit sprache so interessant. URL: <https://www.all-electronics.de/elektronik-entwicklung/deshalb-ist-die-bedienung-mit-sprache-so-interessant-864.html>. Abgerufen am 09.02.2024.
- Sprotte, Y. (2022). Sprach-und stimmanalyse in der wissenschaft: Instrumente der bewertung und des monitorings therapeutischer prozesse. In *Therapeutengeführte Gespräche mit schizophrenen Patienten über ihre eigenen Bilder in der Kunsttherapie: Eine quantitative Einzelfallstudie mittels Text-und Stimmanalyse*, pages 91–123. Springer. DOI: 10.1007/978-3-658-38109-7₅.
- Takayama, H., Takao, T., Masumura, R., Yamaguchi, Y., Yonezawa, R., Sakaguchi, H., Morita, Y., Toyonaga, T., Izumiyama, K., und Kodama, Y. (2023). Speech recognition system generates highly accurate endoscopic reports in clinical practice. *Internal Medicine*, 62(2):153–157. DOI: 10.2169/internalmedicine.9592-22.
- Thakur, A., Ahuja, L., Vashisth, R., und Simon, R. (2023). Nlp & ai speech recognition: an analytical review. In *2023 10th International Conference on Computing for Sustainable Global Development (INDIACom)*, pages 1390–1396. IEEE. ISBN = 978-93-80544-47-2.
- Tscheschner, D. (2007). *Speech and Speaker Recognition*, volume 39. Akademie-Verlag. ISSN =09422919.

- Vogel, M., Kaiser, W., Wassmuth, R., und Mayatepek, E. (2015). Analysis of documentation speed using web-based medical speech recognition technology: randomized controlled trial. *Journal of medical Internet research*, 17(11):e247. DOI: 10.2196/jmir.5072.
- Vogel, M. und Varoß, W.-C. (2024). Microsoft—die rolle von big tech im bereich des health data managements. In *Health Data Management: Schlüsselfaktor für erfolgreiche Krankenhäuser*, pages 823–833. Springer. DOI: 10.1007/978-3-658-43236-2_9.
- Voicedocs (2022). Automatisierte transkription. URL: <https://voicedocs.com/de/blog/spracherkennung-im-gesundheitswesen-diktat-und-transkription>. Abgerufen am 09.04.2024.
- Walker, T., Christensen, H., Mirheidari, B., Swainston, T., Rutten, C., Mayer, I., Blackburn, D., und Reuber, M. (2020). Developing an intelligent virtual agent to stratify people with cognitive complaints: A comparison of human–patient and intelligent virtual agent–patient interaction. *Dementia*, 19(4):1173–1188. DOI: 10.1177/1471301218795238.
- Xiong, X. (2022). A Summary of the Development of Speech Recognition Technology. *Proceedings of the 2022 4th International Conference on Robotics, Intelligent Control and Artificial Intelligence*, pages 768–773. DOI: 10.1145/3584376.3584513.
- Xuedong, H., Baker, J., und Reddy, R. (2014). A historical perspective of speech recognition. *Communications of the ACM*, 57(1):94–103. DOI: 10.1145/2500887.
- Yim, J., Chopra, R., De Fauw, J., und Ledsam, J. (2020). Using artificial intelligence to predict retinal disease progression. URL:<https://deepmind.google/discover/blog/using-ai-to-predict-retinal-disease-progression/>. Abgerufen am 09.04.2024.
- Zuchowski, M. und Göller, A. (2022). Speech recognition for medical documentation: an analysis of time, cost efficiency and acceptance in a clinical setting. *British Journal of Healthcare Management*, 28(1):30–36. DOI: 10.12968/bjhc.2021.0074.

Anhang A

Fragebogen

trifft zu (1), trifft eher zu (2), neutral (3), trifft eher nicht zu (4), trifft nicht zu (5)[Schnell et al., 1999]

1. Zu welchen der nachfolgenden Altersgruppen gehören Sie?
2. Wie leicht konnten Sie sich an die grundlegenden Funktionen gewöhnen?
3. Unterstützt die Benutzeroberfläche eine effektive Arbeitsumgebung?
4. Hat die Verwendung von Dragon zu Verbesserungen in Ihrem Arbeitsauflauf geführt?
5. Empfinden sie die Handhabung des Systems als intuitiv?
6. Erleichtert es die Erfassung umfassender und detaillierter Befunde?
7. Gibt es bestimmte Aspekte der Genauigkeit bei der Worterkennung oder der Erkennung von Abkürzungen, die Ihrer Meinung nach verbessert werden könnten?
8. Wie zufrieden sind Sie mit der Genauigkeit der Spracherkennung und der Fähigkeit des Speechkit, Ihre Sprache korrekt zu transkribieren?
9. Empfinden Sie die Diktierfehler von der Spracherkennung als betriebsverhindernd?
10. Trägt Dragon dazu bei, Fehler in der Befunderfassung zu minimieren?
11. Hat sich die Einführung positiv auf die Dauer der Dokumentation ausgewirkt?
12. Wie zufrieden sind Sie mit der Vielfalt der verfügbaren Textbausteine?
13. Welchen Einfluss hat Dragon auf die Gesamteffizienz Ihrer Arbeitsabläufe?
14. Wie beurteilen Sie die Auswirkungen von Dragon auf die Work-Life-Balance, insbesondere im Hinblick auf die Zeitersparnis oder zusätzliche Belastungen?

15. Wie beurteilen Sie die praktische Relevanz der in den Schulungen vermittelten Informationen?
16. Haben Sie das Gefühl, dass die Schulungen ausreichend auf spezifische Herausforderungen in ihrer Arbeitsumgebung eingegangen sind?
17. Haben die Schulungen Ihre Fähigkeit zu effektiver Nutzung von Dragon verbessert?
18. Bitte teilen Sie Ihre konkreten Gedanken oder Anmerkungen zum Umfang der Schulungsunterlagen(wenn vorhanden) mit und nennen Sie gegebenenfalls weitere Ressourcen, die Ihrer Meinung nach hilfreich wären bzw. Ressourcen, die nicht hilfreich waren.
19. Hat die Anwendung Ihre allgemeine Arbeitszufriedenheit verbessert?
20. Gibt es für Sie mehr Vorteile als Nachteile bei der Benutzung von Spracherkennungstechnologien?
21. Hatten Sie das Gefühl, dass das System mitlernt?
22. Gibt es bestimmte Aspekte von Dragon, die positiv oder negativ hervorheben möchten?
Freitext
23. Können Sie konkrete Beispiele für Situationen oder Herausforderungen nennen, in denen Dragon besonders hilfreich oder nicht hilfreich war? Freitext
24. Haben Sie spezielle Anpassungen vorgenommen, um Dragon besser in Ihren Arbeitsablauf zu integrieren, und wenn ja, welche? Freitext

Anhang B

Interview Transkript

Interviewer: Warum wurde sich für das Dragon Medical Speechkit entschieden?

Interviewpartner: Aufgrund der Unzufriedenheit mit der derzeitigen Diktierlösung (SPEAKING) haben wir uns nach Alternativen umgesehen. Nach einer Marktdurchsichtigung haben wir uns für das Produkt „Dragon Medical One“ entschieden. Das ist jene Lösung, wie sie derzeit noch von den Ärzten in Graz eingesetzt wird. Parallel dazu hat die Firma Care Solutions das Dragon Medical Speechkit sich angeschaut und die ersten Schritte gesetzt, diese Diktierlösung ins KIS zu implementieren.

Interviewer: Wer traf die Entscheidung und wann?

Interviewpartner: Die Entscheidung für Dragon Medical One wurde von IT-BM getroffen und der Nutzungsvertrag wurde im Juni 2021 abgeschlossen, mit einer Mindestnutzungsdauer von 3 Jahren.

Interviewer: Welche Kriterien waren ausschlaggebend?

Interviewpartner: Kriterium war primär die Qualität der Spracherkennung und die Integrationsmöglichkeit ins KIS.

Interviewer: Wurden auch andere Produkte in Betracht gezogen?

Interviewpartner: Der Markt ist hier vom Angebot her gesehen sehr überschaubar. Die Firma Grundig nutzt zum Beispiel auch als Basis das Dragon. Aufgrund der Erfahrungsberichte und selbst geführter Testungen konnte sehr schnell eine Entscheidung getroffen werden.

Interviewer: Wurden spezielle Sicherheitsprotokolle für die Verwendung des Speechkits implementiert?

Interviewpartner: Das Dragon Medical One (DMO) und Dragon Medical Speechkit (DMSK) sind cloudbasierte Dienste. Die URL, die für die Übertragung der Sprachkonserve aufgerufen wird, ist entsprechend verschlüsselt (https).

Interviewer: Wie wurde das Programm problemlos in den vorhandenen Programmworkflow eingefügt?

Interviewpartner: Im Endeffekt funktioniert das Programm wie eine Tastatur. Dort, wo der Cursor gerade positioniert ist, wird der erkannte Text eingefügt. Somit war es auch kein Problem, das Programm in den Programmworkflow einzufügen.

Interviewer: Welche technischen Probleme traten während des Rollouts auf?

Interviewpartner: DMO und DMSK kann nicht in alle KIS-Container den Text übertragen, da es sogenannte „Window Classes“, z. B. RadRichTextEditor, gibt, die vom Produkt nicht unterstützt werden. Hier gibt es einen Workaround, der jedoch dann andere Nachteile hat, wie z. B. dass das sprachgesteuerte Korrigieren mittels Sprachbefehlen nicht mehr möglich ist.

Interviewer: Wie gut integriert sich die neue Spracherkennungstechnologie mit den bereits vorhandenen klinischen Informationssystemen?

Interviewpartner: Das DMO und sind im klinischen Informationssystem sehr gut integriert. Zukünftig wird es jedoch Erweiterungen geben, wie z. B. der sprachgesteuerte Aufruf von bestimmten Sichten, wie z. B. die Fieberkurve oder „Öffne Vitalparameter Blutdruck“, und es öffnet sich dann ein Fenster, wo ich diesen erfassen kann mittels Diktieren.

Interviewer: Gab es technische oder operative Hürden bei der Integration, und wie wurden diese überwunden?

Interviewpartner: Ausgehend von Punkt 5 kann Speechkit hier Abhilfe schaffen und die Diktiermöglichkeit auch in „Window Classes“ schaffen, die out-of-the-box von DMO nicht unterstützt werden, inklusive der Möglichkeit, z. B. sprachgesteuert korrigieren zu können. Beim Speechkit selbst gab es keine Hürden bei der Integration.

Anhang C

Rohdaten Fragebogen

Anhang C Rohdaten Fragebogen

Können Sie konkrete Beispiele für Situationen oder Herausforderungen nennen, in denen Dragon besonders hilfreich oder nicht hilfreich war?	Haben Sie spezielle Anpassungen vorgenommen, um Dragon besser in Ihren Arbeitsablauf zu integrieren und wenn ja, welche?
-	-
.	.
Hilfreich: bei Aufnahmegesprächen ist es sehr hilfreich, da Pat. Gleich hören, was man einschreibt	Ja, Umstrukturierung von Aufnahmeprozess
viele Endoskopien, viele schnelle Übernahmen möglich	nein
.	nein
nein	nein
.	.
.	.
vlt in Notsituationen	leider noch nicht
.	.
notsituationen	noch nicht
noch zu wenig benutzt	unstrukturieren bei der Dokumentation
nachtdienst	regelmäßiges Ausprobieren
bei lauten Hintergrundgeräuschen geht es nicht gut	nein
-	.
nachtdienst	.
bei nur Benutzung von Textbausteine große Zeitersparnis	diktieren in einem ruhigen Raum
-	.
Dokumentation in Nachtdiensten bei längeren Einträgen (Aufnahme)	Nein
-	-
funktioniert nicht immer im Stress	ich verwende es immer
Pflegebericht	Nein
Bei Aufnahmegespräche sowie lange Pflegeberichte sehr hilfreich	Nein
Rechtschreibung!	Nein
im Nachtdienst	.
-	-
.	.
wenn man nur schnell etwas einfügen muss	nein
.	.
.	.
-	nein
.	nein

Gibt es für Sie mehr Vorteile als Nachteile bei der Benutzung von Spracherkennungstechnologien?	Hat die Anwendung Ihre allgemeine Arbeitszufriedenheit verbessert?	Haben Sie das Gefühl, dass das System mitlernt?
4	3	-
1	1	.
2	1	Positiv: Zeitersparnis, Negativ: jeder kann mithören ;)
1	1	nein
2	1	Lernt mit
2	2	nein
2	1	Negativ: im Tagesablauf schwierig einzusetzen aufgrund von lauten Hintergrundgeräuschen aufgrund des Tagesbetriebs. Im Nachtdienst bei Ruhe leichter zu benutzen.
4	3	.
3	3	ist nicht zeitsparend für mich
3	3	Textbausteine sehr hilfreich!!
2	1	Positiv: Textbausteine, negativ: Maus-Cursor
2	2	Abkürzungen
3	2	weiteres technisches Gerät, sperrig zum Mitnehmen bei Laptops
3	2	Textbausteine, keine Schreibfehler
3	1	Abkürzung Textbausteine
3	2	Personalisierbares Vokabular
3	1	Textbausteine
2	1	Textbausteine
5	2	Nein
2	3	-
1	1	Praktisch
2	2	Performance
2	3	Spracherkennung funktioniert gut
1	1	Kabelloses Gerät wäre besser, oder Headset, einfacheres Gerät mit nur Start- und Stoppknopf
3	2	Textbausteine sind nützlich aber der zusätzliche Kabel nervt
2	1	Username-Wechsel
3	2	Textbausteine
3	5	Zeitverzögerung, Erkennung von Textbausteinen
1	2	Positiv: Textbausteine, negativ: erkennt Wörter falsch
1	1	.
1	2	Weiteres technisches Kästchen, Maus springt, wenn man das Mikro falsch hält oder sozusagen auf den Bauch legt
3	5	Positiv: Zeitersparnis Dokumentation

Anhang C Rohdaten Fragebogen

Wie beurteilen Sie die praktische Relevanz der in den Schulungen vermittelten Informationen?	Haben Sie das Gefühl, dass die Schulungen ausreichend auf spezifische Herausforderungen in Ihrer Arbeitsumgebung eingegangen sind?	Haben die Schulungen Ihre Fähigkeit zu effektiver Nutzung von Dragon verbessert?	Bitte teilen Sie Ihre Konkreten Gedanken oder Anmerkungen zum Umfang der Schulungsunterlagen (wenn vorhanden) mit und nennen Sie gegebenenfalls weitere Ressourcen, die Ihrer Meinung nach hilfreich wären?	Hat die Anwendung Ihre allgemeine Arbeitszufriedenheit verbessert?
4	3	4	-	4
1	1	2	.	1
2	3	2	.	1
1	1	1	nichts gefunden	1
1	2	1	1	.
1	1	1	Hatte keine Schulung	2
1	1	2	.	1
2	1	1	.	4
2	2	1	Problem ist, dass ich beim Schreiben viel rascher bin	1
4	4	3	.	4
2	2	1	.	2
2	2	1	mehr Textbausteine	2
2	2	2	hatte keine Schulung	1
1	1	1	.	2
1	2	1	.	2
2	2	2	keine Schulungsunterlagen erhalten	2
2	2	1	Textbausteine sehr hilfreich; gerne mehr Übernahmen (z. B. Kolo)	2
1	2	2	.	2
3	3	3	.	1
1	1	1	-	2
1	1	1	Schulung war top	1
2	2	2	Leider keine vorhanden	2
2	3	3	Textbausteine sind ausbaufähig	2
1	1	2	Austeilen einer Kurzbeschreibung, da der SpeechMike viele Tasten hat, auch wenn man sie nicht benötigt. Austeilen der Textbausteine.	1
2	2	1	Bereitstellung von der Textbausteinliste sehr nützlich	2
2	2	1	.	2
2	1	2	hätte gerne mehr bekommen	3
2	1	4	Kollegen haben geholfen	3
2	2	1	Textbausteine sehr hilfreich, aber ein Heftchen, wo die Fehlermeldungen drinnen stehen	2
2	2	1	Textbausteine sehr hilfreich	2
2	1	1	Textbausteine hilfreich	2

Anhang C Rohdaten Fragebogen

Wie zufrieden sind Sie mit der Vielfalt der verfügbaren Textbausteine?	Welchen Einfluss hat das SpeechKit auf die Gesamteffizienz Ihrer Arbeitsabläufe?	Wie beurteilen Sie die Auswirkungen von Dragon auf die Work-Life Balance, insbesondere im Hinblick auf die Zeitersparnis oder zusätzliche Belastung?
2	nicht viel	-
2	Erleichterung	-
2	Nach einer Eingewöhnungsphase ist man mit der Benutzung von Dragon auf jeden Fall schneller	Keine Auswirkungen
1	wird flüssiger, der Arbeitsablauf, je länger man damit arbeitet	sehr gut
1	als Nachsorgeprotokoll zum Eintragen ist eine Erleichterung	gut
2	Man ist schneller dadurch	Mehr Zeit für die Arbeit am Patienten
1	Im Nachtdienst deutliche Verbesserung der Schnelligkeit der Dokumentation.	-
4	keine, da ich es nicht anwende - und mit dem schreiben schneller bin	Dauer der Dokumentation wird verlängert
4	noch keine effiziente Antwort zu geben	für mich nicht relevant
3	Dokumentationsprozess im gesamten schneller, durch die Eingewöhnung aber langwieriger	-
2	effizienter besonders Textbausteine	ingsamt kürzer aber eingewöhnung schwierig
2	neue gewöhnung	tippen geht derzeit noch schneller
1	-	zeitersparnis
3	geht schneller	zusätzliche Belastung in der Eingewöhnungsphase
1	-	-
2	spart insgesamt zeit aber am Anfang mehr aufwand	-
2	schneller	zusätzliche Belastung
+2	schneller	-
2	Leichte Verbesserung	Es wurden keine signifikanten Unterschiede vor oder nach der Benutzung festgestellt
3	meist schnellere Dokumentation	-
3	Geht alles schneller	super
1	Einen großen	Als Zeitersparnis
3	Schnellere Dokumentation im Nachtdienst	?
2	Gute Struktur, Hemmschwelle, des Redens vor den anderen Mitarbeiter, muss überwunden werden.	Zeitersparnis ist definitiv gegeben.
2	keinen	am Anfang zusätzliche Belastung
2	sonst längere dokumentation geht schneller, man kann längere Berichte in kurzer Zeit schreiben	beim Pflegestützpunkt ist es schwierig in ruhe zu diktieren
2	dauert länger als tippen	-
2	momentan eher ein hindernis	zusätzlicher zeitaufwand im arbeitsalltag
2	dokumentationsprozess ist etwas komplexer	am Anfang zusätzliche Belastung
2	schnellere Dokumentatiomm von langen texten	umgewöhnung bedeutet zusätzliche belastung
2	am anfang verlangsamerung, jetzt gehts schneller, auch mithilfe personalisierter Abkürzungen	-

Anhang C Rohdaten Fragebogen

Gibt es bestimmte Aspekte der Genauigkeit bei der Worterkennung oder der Erkennung von Abkürzungen, die Ihrer Meinung nach verbessert werden könnten?	Wie zufrieden sind Sie mit der Genauigkeit der Spracherkennung und der Fähigkeit des SpeechKit, Ihrer Sprache korrekt zu transkribieren?	Empfinden Sie die Diktiertfehler von Spracherkennung als betriebsverhindernd?	Trägt Dragon dazu bei Fehler in der Befurfassung zu minimieren?	Hat sich die Einführung positiv auf die Dauer der Dokumentation ausgewirkt?
Abkürzungen	2	4	3	2
.	1	4	2	1
Ja	1	2	1	1
man muss sehr deutlich sprechen	2	3	2	1
Namen der Patienten wurde nicht immer richtig verstanden	2	2	3	1
weiß nicht	2	4	3	3
Teilweise Spracherkennungsprobleme bei ähnlich klingenden Wörtern wie Garten oder Gatte	1	3	3	2
.	4	3	2	3
Bauklötze	2	3	3	3
einigen Abkürzungen wie zB "LAS", "lt. Pat." werden nicht oder falsch erkannt	4	3	3	2
bei manche Abkürzungen und Wörter Erkennung falsch	2	4	3	2
abkürzungen.herzrythmus	2	4	2	1
abkürzungen, Sprechstil bei Textbausteinen	2	4	2	2
Abkürzungen falsch erkannt	2	4	2	1
Abkürzungen	2	4	2	1
abkürzungen	2	4	2	2
die benötigte Sprechgeschwindigkeit bei den Textbausteinen	2	4	3	2
Erkennung von Abkürzungen	2	5	2	2
Nein	2	5	4	3
.	3	3	3	2
Befehle wie nächste Zeile, Punkt etc.	2	2	1	1
Nein	2	4	3	2
Nein	2	3	3	2
Wird alles nach erlernen der Worte gut erkannt	1	2	1	2
teilweise falsche Erkennung aber nach speziellen erlernen transkribiert er richtig	2	4	1	2
-	2	4	1	2
Textbausteine werden nicht erkannt	4	3	2	4
Erkennung von Worten funktioniert zu oft nicht	4	1	5	4
mikrofon ohne kabel wäre besser	2	4	2	1
-	2	4	2	1
mehr Textbausteine	2	4	2	1

Anhang C Rohdaten Fragebogen

id	Zu welcher der nachfolgenden Altersgruppe gehören Sie?	Konnten Sie sich leicht an die grundlegenden Funktionen der Spracherkennung gewöhnen?	Sie sich an die grundlegenden Funktionen der Spracherkennung gewöhnen?	Unterstützt die Benutzeroberfläche einen effektiven Arbeitsablauf?	Hat die Verwendung der Spracherkennung zu Verbesserungen in Ihrem Arbeitsablauf geführt?	Empfinden Sie die Handhabung des Systems als intuitiv?	Erleichtert es die Erfassung umfassender und detaillierter Befunde?
1	21-29	4	3	4	4	4	4
2	40-49	1	2	1	1	2	1
3	30-39	1	2	1	1	1	2
4	50-59	1	2	1	1	1	1
5	30-39	1	2	2	2	2	2
6	30-39	2	2	2	3	2	3
7	21-29	2	2	2	3	1	1
8	21-29	1	1	1	3	4	4
9	50-59	3	3	3	4	2	2
10	21-29	5	4	4	3	2	4
11	30-39	1	2	2	2	1	1
12	40-49	1	2	2	4	1	3
13	21-29	2	2	2	2	1	2
14	30-39	2	2	2	1	1	1
15	40-49	2	2	2	1	2	2
16	30-39	2	2	2	2	1	1
17	21-29	4	3	3	2	2	3
18	40-49	1	2	1	1	1	2
19	21-29	1	2	2	4	1	3
20	21-29	1	2	2	2	2	3
21	21-29	2	2	2	1	2	1
22	30-39	1	2	2	2	2	2
23	30-39	1	2	2	2	3	3
24	40-49	1	1	1	1	2	1
25	21-29	2	2	2	1	2	1
26	30-39	1	2	2	3	1	2
27	50-59	3	2	2	3	2	1
28	30-39	3	4	4	4	2	4
29	40-49	2	1	1	2	1	2
30	21-29	2	1	1	3	1	2
31	21-29	2	1	1	2	1	1
32	21-29	1	2	2	3	2	2

Eidesstattliche Erklärung

zur Arbeit mit dem Titel

DRAGON: ERFASSUNG VON BEFUNDEN MITTELS SPRACHERKENNUNG

“Ich erkläre hiermit an Eides statt,

- dass ich die vorliegende Bachelorarbeit/Masterarbeit selbstständig angefertigt und die mit ihr verbundenen Tätigkeiten selbst erbracht habe und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe;
- dass ich mich bei der Erstellung der Arbeit an die Richtlinie der FH JOANNEUM zur Sicherung Guter Wissenschaftlicher Praxis und zur Vermeidung von Fehlverhalten (kurz Richtlinie GWP) gehalten habe;
- dass ich alle aus gedruckten oder ungedruckten Werken sowie aus dem Internet im Wortlaut oder im wesentlichen Inhalt übernommenen Formulierungen und Konzepte gemäß den Regeln für gutes wissenschaftliches Arbeiten (Richtlinie GWP) zitiert und durch genaue Quellenangaben gekennzeichnet habe;
- dass ich in der Methodendarstellung oder einem Verzeichnis alle verwendeten Hilfsmittel (Assistenzsysteme der Künstlichen Intelligenz wie Chatbots [z.B. ChatGPT], Übersetzungsapplikationen [z.B. DeepL], Paraphrasierapplikationen [z.B. Quill bot]), Bildgeneratorapplikationen [z.B. Dall-E] oder Programmierapplikationen [z.B. Github Copilot] deklariert und ihre Verwendung bei den entsprechenden Textstellen angegeben habe;
- dass die vorliegende Originalarbeit in dieser Form zur Erreichung eines akademischen Grades noch keiner anderen Hochschule vorgelegt worden ist. (* Diese Formulierung entfällt bei gemeinsamen Studienprogrammen, z.B. Double Degree Programmen. Es gelten die hierfür festgelegten Regelungen.)

Ich wurde darüber aufgeklärt, dass meine Arbeit auf Plagiate und auf Drittautor:innenschaft menschlichen (Ghostwriting) oder technischen Ursprungs (Assistenzsysteme der künstlichen Intelligenz) überprüft werden kann.

Ich bin mir darüber im Klaren, dass eine wahrheitswidrige Erklärung rechtliche Folgen wie eine negative Beurteilung meiner Arbeit, die nachträgliche Aberkennung des dadurch erlangten Titels und Strafverfolgung nach sich ziehen kann.”