



Hanna Ehlert & Ulrike Lüdtko

Künstliche Intelligenz (KI) in der Logopädie

Nutzungsrechte

Potenzial und Herausforderungen am Beispiel einer Automatisierung der Diagnostik im Bereich Kindersprache

Einleitung

Künstliche Intelligenz (KI) ist längst in unserem Alltag angekommen. Wir nutzen die Gesichtserkennung, um unser Handy zu entsperren, wir fragen die Software Siri, um ein indisches Restaurant in der Nähe zu finden, wir suchen eine*n Partner*in per Dating-App und erhalten personalisierte Werbung, wenn wir den Computer benutzen. Alle diese Systeme basieren auf KI. Eine einheitliche Definition des Begriffs existiert allerdings nicht. Die meisten beschreiben KI als Nachbildung von Leistungen menschlicher Kognition mittels Computersoftware (z.B. *Zandi et al.* 2019). Der Philosoph *John Searle* (1980) eta-

blierte die Unterscheidung zwischen „starker“ und „schwacher“ KI. Unter „schwacher“ KI versteht man eine Software, die in der Lage ist, große Datenmengen für einen bestimmten Zweck zu verarbeiten. Nur in ihrem begrenzten Anwendungsbereich (z.B. Schach) ist „schwache“ KI in der Lage, menschliche Leistungen zu übertreffen. „Starke“ KI hingegen hat tatsächlich mehr Ähnlichkeit mit menschlicher kognitiver bzw. mentaler Verarbeitung. Sie erlangt ein tieferes Verständnis für die ihr vorgelegten Probleme und deren Lösung, sie kann Kausalitäten erkennen, logisch und empathisch denken sowie bewusst handeln. Aktuell sind alle KI-Anwendungen der Gruppe der „schwachen“ KI zu-

zuordnen (*Taulli* 2022). Dennoch nimmt der Einfluss, den KI in allen Lebensbereichen hat, stetig zu. Auch im Gesundheitswesen wird ihr Einsatz vor allem in Form des Maschinellen Lernens erforscht. Chancen, die in diesem Zusammenhang benannt werden, liegen in der Möglichkeit, große Mengen von Gesundheitsdaten zu verarbeiten und so frühzeitiger und präzisere Diagnosen zu stellen, Abläufe im Gesundheitswesen zu optimieren und dadurch Ressourcen einzusparen sowie menschlichen Bias in der medizinischen Entscheidungsfindung zu vermindern (*Murphy et al.* 2021, *Yu et al.* 2018). Um diese Entwicklung im Sinne der Anwendungen gestalten zu können, ist es notwendig, sich mit KI auseinanderzusetzen. Der Artikel fokussiert deshalb exemplarisch das für die Logopädie relevante Maschinelle Lernen als Teilgebiet der Künstlichen Intelligenz und als Anwendungsbereich die Kindersprachdiagnostik.

ZUSAMMENFASSUNG. Dieser Artikel bietet eine Einführung und einen Überblick zur KI-basierten Automatisierung in der Logopädie am Beispiel der Diagnostik im Bereich Kindersprache. Die Funktionsweise Künstlicher Intelligenz und spezifischer des Maschinellen Lernens wird anhand der relevanten Anwendungen der automatischen Spracherkennung und automatischen Textverarbeitung eingeführt. Hierbei werden aktuelle Möglichkeiten, aber auch Grenzen dieser Technologie mittels ausgewählter diagnostischer Themenfelder (Automatisierung in der Diagnostik von Aussprachestörungen, der Auswertung von expressiven Sprachentwicklungstests und der Spontansprachanalyse) vorgestellt. Abschließend werden technische und ethische Herausforderungen von KI-basierter Software für die Logopädie verdeutlicht und Bedarfe für einen gewinnbringenden und sinnvollen Umgang unserer Berufsgruppe mit dieser Technologie abgeleitet.

SCHLÜSSELWÖRTER: Künstliche Intelligenz (KI) – Maschinelles Lernen – Diagnostik – Kindersprache – automatische Spracherkennung – automatische Textverarbeitung

Maschinelles Lernen

Maschinelles Lernen versetzt Computer in die Lage, ihre Leistungsfähigkeit eigenständig durch die Verarbeitung von Daten zu verbessern. Hierin besteht der größte Unterschied zu einer Software wie computerbasierte Therapieprogramme oder Testverfahren, die im Rahmen der Digitalisierung in der logopädischen Praxis bereits genutzt werden. Herkömmliche Software basiert auf AI-

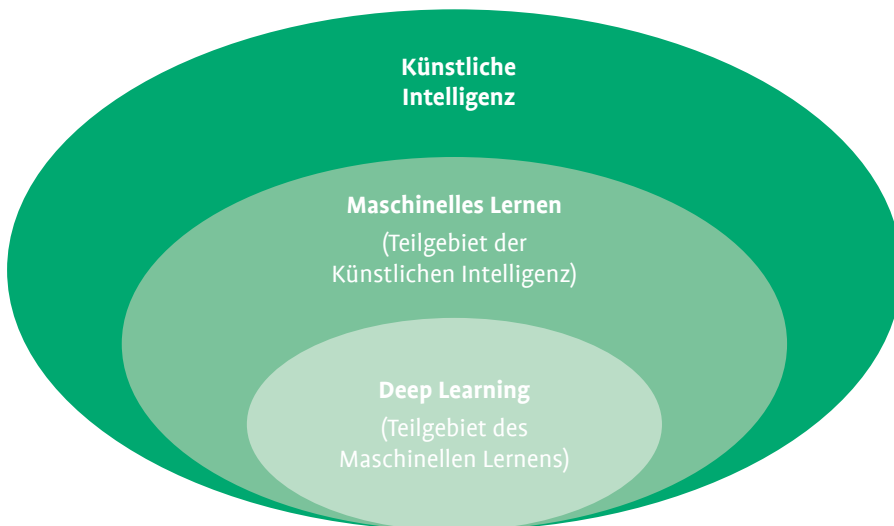
gorithmen, denen eine eindeutige Handlungsvorschrift zur Lösung eines Problems oder zur Ausführung einer Aufgabe einprogrammiert ist. Diese Software ist deshalb in ihrem Einsatz stark begrenzt (z.B. Eingaben addieren, Videos abspielen), sie kann weder komplexen Input (z.B. Sprache) verarbeiten noch flexibel auf solchen Input reagieren. Beim Maschinellen Lernen hingegen erarbeiten Computersysteme selbstständig Lösungen für eine definierte Aufgabe anhand von Beispielen (Daten) und bestimmten Vorgaben (Regeln). Die Hauptanwendungsbereiche solcher Systeme sind die Mustererkennung und darauf basierende Vorhersagen (Taulli 2022). Ihre Entwicklung folgt dabei unabhängig vom Anwendungsbereich immer demselben Prinzip: Zunächst werden annotierte (gekennzeichnete) Daten mit bekannten Zusammenhängen benötigt, anhand derer ein Modell für die Mustererkennung trainiert wird. In einem zweiten Schritt wird diese Fähigkeit auf nicht-annotierte, unbekannte Daten übertragen (Ertel 2021). Soll ein Modell etwa Gesichter erkennen, werden als Trainingsdaten Bilder von Menschen benötigt, auf denen die Gesichter markiert sind. Hat das Modell genügend solcher Bilder mit markierten Gesichtern verarbeitet, kann es sie auch auf neuen Bildern erkennen, auf denen die Gesichter nicht markiert sind. Eine heutzutage häufig zum Einsatz kommende Variante des Maschinellen Lernens sind neuronale Netzwerke. Analog zu den Neuronen und Synapsen des menschlichen Gehirns bestehen künstliche neuronale Netze aus vielen Schichten von Datenknoten, die durch gewichtete Verbindungen vernetzt sind. Die Gewichtung der einzelnen Verbindungen zwischen den Schichten wird in der Entwicklung des Netzwerks immer wieder angepasst, um die Leistung des Mo-

dells zu optimieren. Solche Modelle werden „Deep Neural Networks“ oder „end-2-end“ genannt. Dadurch können beim sogenannten „Deep Learning“ immer komplexere Probleme gelöst werden (Ertel 2021). Abbildung 1 verdeutlicht die – auch als historische Entwicklung zu verstehende – Beziehung der Begriffe „Künstliche Intelligenz“, „Maschinelles Lernen“ und „Deep Learning“ zueinander.

Anwendungen in der logopädischen Diagnostik

Um die Funktionsweise von Anwendungen Maschinellen Lernens zu verstehen, die für den logopädischen Praxiskontext relevant sind, ist eine datenbasierte Perspektive notwendig. So betrachtet, liegt menschliche Kommunikation entweder als Audiodaten (gesprochene Sprache) oder Textdaten (verschriftete gesprochene Sprache/Schriftsprache) vor. Daher wird eine Software benötigt, die Audiodaten verarbeitet und analysieren sowie verbale Audiodaten in Textdaten umwandeln und diese Textdaten wiederum verarbeiten und analysieren kann. Programme, die verbale Audiodaten verarbeiten (und analysieren) sowie verbale Audiodaten in Textdaten konvertieren, werden als automatische Spracherkennung (automatic speech recognition, ASR) bezeichnet; Programme, die Textdaten verarbeiten (und analysieren), als automatische Textverarbeitung (natural language processing, NLP). Dabei ist anzumerken, dass die englischen Begrifflichkeiten nicht immer trennscharf verwendet werden, da unterschiedliche Disziplinen (u.a. Ingenieurwissenschaften, Computerlinguistik) an der Entwicklung der entsprechenden Programme arbeiten.

Abb. 1: Künstliche Intelligenz und Teilgebiete



Automatische Spracherkennung (ASR)

ASR ist der Prozess der Konvertierung von gesprochener Sprache in Text. Das zugrunde liegende mathematische Modell erhält als Input Audiodaten (oft auch weiterverarbeitet als Spektrogramm) und gibt als Output eine Sequenz von Zeichen oder Phonemen (z.B. als Transkript) aus (Alharbi et al. 2021). Für die Erstellung des Outputs wird in manchen Modellen sprachspezifisches Wissen (Übergangswahrscheinlichkeiten zwischen Lauten und Wörtern) genutzt (sog. Hidden-Markov-Modelle, HMM), viele arbeiten rein datenbasiert (End-2-End-Systeme, Deep Learning) (Keshet 2018). In jedem Fall findet im Rahmen der Verarbeitung des Inputs eine Segmentierung statt. Trainiert wird das Modell durch Audios und zugehörige manuell erstellte Transkripte. Für die Leistungsfähigkeit eines ASR-Modells sind folgende Faktoren entscheidend:

- Die Menge an Trainingsdaten, die für die Entwicklung zur Verfügung stehen. Je mehr qualitativ hochwertige Daten, desto besser.
- Die Ähnlichkeit der Trainingsdaten mit der Sprache, für die die Software später angewendet werden soll. Ein Modell, das auf Erwachsenensprache trainiert wurde, funktioniert nicht fehlerfrei für Kindersprache.
- Die Größe des Vokabulars des Modells. ASR-Software variiert in Bezug auf die Menge an Wörtern, die sie erkennen kann.
- Die Art der Sprache, die erkannt und verarbeitet werden soll. Typische Sprache ist leichter zu verarbeiten als auffällige Sprache.
- Der akustische Kontext, in den die zu erkennende Sprache eingebettet ist. Sprache, die im Studio aufgenommen wurde, ist einfacher zu verarbeiten als Sprache, die zusammen mit Hintergrundgeräuschen aufgenommen wurde (Alharbi et al. 2021).

Automatische Textverarbeitung (NLP)

NLP meint Programme, die schriftliche Texte in interpretierbare Datensets konvertieren, die wiederum mithilfe statistischer Modelle und Modellen des Maschinellen Lernens analysiert werden können (Harrison & Sidey-Gibbons 2021). Diese Programme, die ihren Ursprung an der Schnittstelle zwischen Künstlicher Intelligenz und Linguistik haben, können große Mengen an Textdaten strukturieren und segmentieren, um daraus spe-

zifische Informationen zu extrahieren (z.B. grammatische Strukturen). Die zentrale Aufgabe für diese Systeme besteht darin, Texte aus natürlicher Kommunikation in eindeutige Computerrepräsentationen zu übersetzen, die für programminterne Zuordnungen und Abfragen geeignet sind (Nadkarni et al. 2011). Automatische Textverarbeitung arbeitet hierarchisch meist von kleineren zu größeren textlichen Einheiten. Üblicherweise wird zunächst auf Wortebene die morphologische Struktur bestimmt, auf Satzebene die Reihenfolge der Wörter, Syntax und Bedeutung des gesamten Satzes, um schließlich den Kontext des gesamten Textes zu analysieren. Dabei erfolgt zunächst eine automatische Wort- bzw. Tokenerkennung (tokenization), welche den Text in kleinere Segmente zerlegt. Weitere häufig angewandte – meist sprachenabhängige – Annotationstechniken der automatischen Textverarbeitung sind:

- Bestimmung von Satzgrenzen, Satzanfängen und -enden
- Zerlegen eines Satzes in seine grammatischen Bestandteile (parsing)
- Reduzierung der flektierten Formen eines Wortes auf eine einzige, verarbeitbare Form (lemmatization)
- Identifikation von Wortstämmen (stemming)
- Klassifizierung und Benennung von Wortteilen (Part-of-speech-Tagging/ POS-Tagging)

Nachdem ein Text mit den genannten Techniken bearbeitet wurde, kann er über Abfragen und Kombination von Informationen linguistisch analysiert werden. Die eben beschriebenen Abläufe des NLP sind dem klassischen Maschinellen Lernen zuzuordnen, aktuelle Entwicklungen wie etwa GPT (Generative Pre-Trained Transformer) basieren demgegenüber – wie auch die meisten ASR-Anwendungen – auf End-2-End-Modellen, also Deep Learning (Topal et al. 2021).

Forschungsstand zum Einsatz Maschinellen Lernens in der Diagnostik von Kindersprache

Um einen Eindruck über die aktuellen Möglichkeiten und Grenzen des Einsatzes von ASR und NLP in der Logopädie zu erhalten, folgt ein Überblick zur Forschung und bereits eingesetzten Tools. Audiodaten werden in der automatisierten Diagnostik genutzt, wenn rein akustische Phänomene betrachtet werden (z.B. Betonung, Sprechflüssigkeit) und Textdaten, wenn linguistische Phäno-

me ne betrachtet werden (z.B. phonologische Prozesse, grammatikalische Korrektheit).

Automatisierung der Diagnostik von Aussprachestörungen

Kindliches Sprechen wird hauptsächlich bei Aussprachestörungen wie phonologischen Störungen oder kindlicher Sprechapraxie evaluiert. In einem Review von McKechnie et al. (2018) werden für den Zeitraum 2007 bis 2016 insgesamt 32 Studien identifiziert, die sich mit dem Thema der automatisierten Sprechdiagnostik bei Kindern beschäftigen. Dieser Bereich hat bisher die meiste Forschung generiert.

Ein Fokus lag dabei auf der automatisierten Diagnostik von Kindern mit und ohne Sprechapraxie, wohl auch, weil entsprechende Symptome hier gut von unbeeinträchtigter Aussprache zu trennen sind. Zur Klassifizierung der Kinder mit Sprechapraxie wurde deren Aussprache auf Einzelwortebene (teils Realwörter, teils Pseudowörter) analysiert und typische Auffälligkeiten z.B. in der Betonung berücksichtigt (McKechnie et al. 2021, Shahin et al. 2015, Shulga et al. 2020).

Die Bedeutsamkeit der Menge an Trainingsdaten verdeutlichen zwei Studien, die akzeptable Ergebnisse (> 80%) in der automatisierten Identifikation von typischen phonologischen Prozessen im Chinesischen erreichen konnten (Chen 2011, Ng & Lee 2020). In beiden Studien wurden die genutzten Modelle mit umfangreichen Datensätzen trainiert: Audioaufnahmen der Einzelwortproduktion von über 400 Kindern bei Chen (2011) und von fast 2.000 Kindern bei Ng und Lee (2020).

Automatisierung der Auswertung von expressiven Sprachentwicklungstests

Die KI-basierte Auswertung eines expressiven Sprachtests wurde bisher nur von der Forschungsgruppe um Gale et al. (2020, 2021)

erprobt. In deren Studien wurden expressive Sprachtests (u.a. expressiver Wortschatz/Benenntest, Nachsprechen von Sätzen, Satzproduktion nach Bildstimulus) durchgeführt und die Antworten manuell transkribiert. Die Auswertung erfolgte dann im Vergleich manuell und automatisiert mittels NLP. In den Ergebnissen zeigte sich, dass die automatische Auswertung signifikant mit der manuellen korrelierte ($p = 0.76-0.99$, $p < 0.001$). Die größten Übereinstimmungen wurden auf Wortebene (Benenntest Wortschatz), die geringsten auf Satzebene (Satzproduktion) erzielt. In einem nächsten Schritt soll ein ASR-System in das Modell integriert werden, um die manuelle Verschriftlichung der Antworten zu ersetzen.

Automatisierung von Spontansprachanalysen

Die technischen Anforderungen an eine ASR steigen mit dem Freiheitsgrad und der Variabilität der zu erkennenden Sprache. So ist ASR auf Wortebene weniger komplex als auf Satzebene, wie das eben genannte Beispiel zeigt. Die größte Herausforderung, aber auch die größte Arbeitserleichterung in der Praxis, stellt die Automatisierung der Spontansprachanalyse dar.

Spontansprachanalysen sind eine anerkannte Methode der Kindersprachdiagnostik, die wertvolle Informationen zu den realsprachlichen Fähigkeiten von Kindern liefert. Aufgrund des hohen Zeitaufwands wird sie jedoch nicht routinemäßig in der logopädischen Praxis eingesetzt (Klatte et al. 2021). Digitale Tools, wie Computerized Language ANalysis (CLAN, MacWhinney 2000), Systematic Analysis of Language Transcripts (SALT, Miller & Chapman 1985) und Language Environment Analysis (LENA™, Gilkerson & Richards 2008) existieren zwar, werden aber eher in der Forschung eingesetzt, weil sie keine vollständige Automatisierung aller Prozessschritte der Spontansprachanalyse bieten (Lüdtke et al. 2023).

Tab. 1: Software zur automatischen Spontansprachanalyse im Vergleich

	Aufnahme der Sprachproben	Transkription der Sprachproben	Annotation der Sprachproben	Analyse der Sprachproben
CLAN	keine Hardware enthalten	manuell	teilweise automatisch (NLP)	textbasiert (NLP): distributiv und linguistisch
SALT	keine Hardware enthalten	manuell	manuell	textbasiert (NLP): distributiv und linguistisch
LENA™	Hardware für bis zu 16 Std. Aufnahmen im Alltag	wird nicht erstellt	wird nicht erstellt	audiobasiert (ASR): distributiv, zusätzlich akustische Umgebung

CLAN und SALT benötigen ein manuelles Transkript der Sprachprobe im programm-eigenen Format als Eingabe und ermöglichen dann eine umfangreiche automatisierte (teilweise unter Einsatz von NLP-Techniken) linguistische Analyse, schwerpunktmäßig auf den Ebenen Wortschatz und Grammatik.

LENA™ setzt sich aus einem kleinen Aufnahmegerät, welches das Kind am Körper trägt und so mehrere Stunden Alltagskommunikation aufgenommen werden können, und einer zugehörigen Auswertungssoftware zusammen. Das Tool überspringt eine Transkription der Sprachproben und liefert eine auf ASR-basierende automatische Auswertung der Audioaufnahme. Allerdings können deshalb nur distributive Parameter (Verteilungsparameter) generiert werden, die keinen Zugang zum linguistischen Inhalt des Gesagten erfordern, wie etwa die Anzahl der Wörter und Äußerungen und ihre zeitliche Dauer über den Tag verteilt. Zusätzlich liefert das Tool eine Auswertung der akustischen Umgebung (z.B. die Dauer, welche das Kind über den Tag mit Medienkonsum verbracht hat). Tabelle 1 stellt die drei Tools vergleichend gegenüber.

Ein Beispiel für die Erforschung der vollständigen Automatisierung von Spontansprachanalysen in Deutschland ist das interdisziplinäre TALC-Projekt (Tools for Analyzing Language and Communication) an der Leibniz Universität Hannover.

Zusammenfassend bietet eine KI-basierte Automatisierung in der Kindersprachdiagnostik Potenzial zur Arbeitserleichterung und Erweiterung der diagnostischen Möglichkeiten. Erste vielversprechende Ergebnisse für den Einsatz von ASR liegen vor. Die diagnostische Klassifikationsfähigkeit der Programme erreicht aber aktuell noch keine klinische Bedeutsamkeit, diese wäre die Voraussetzung für einen flächendeckenden Einsatz in der Praxis.

Herausforderungen des Einsatzes Maschinellen Lernens in der Logopädie

Technische Herausforderungen

Automatische Spracherkennung für Erwachsenensprache ohne Auffälligkeiten und deutliche Dialekte funktioniert bereits mit einer Genauigkeit, die sich der Qualität manueller Transkription annähert (Radford et al. 2022). Die technischen Herausforderungen beziehen sich vor allem auf bestimmte Personengruppen wie jüngere Kinder, Sprecher*innen mit starkem Dialekt, mehrsprachige Sprecher*innen und solche mit sprach-

lichen oder stimmlichen Auffälligkeiten. Eine weitere Herausforderung sind variable (natürlichere) Aufnahmekontexte. Der Schlüssel ist die Menge an zur Verfügung stehenden repräsentativen Trainingsdaten: Je größer die Heterogenität der zu verarbeitenden Zieldaten, desto größer die benötigte Menge an Trainingsdaten. Besonders im Bereich Kindersprache herrscht ein Mangel an öffentlich zugänglichen und für das Maschinelle Lernen aufbereiteten Daten.

Ethisch-rechtliche Herausforderungen

Mit dem Einsatz von KI im Gesundheitswesen sind neben dem vielfältigen Potenzial auch große Herausforderungen verbunden. In der Literatur werden unterschiedliche ethisch-rechtliche Aspekte thematisiert, vier davon werden besonders häufig diskutiert: Fairness, Explainability, Accountability und Privacy.

Fairness meint im Kontext von Maschinellern Lernen die Vermeidung von Diskriminierung und Bias durch Modelle und Trainingsdaten. Denn es besteht die Gefahr, dass sich Annahmen (z.B. Vorurteile) der Entwickler*innen in den Algorithmen wiederfinden. Auch nicht repräsentative, veraltete, ungenaue oder unvollständige Trainingsdaten verhindern eine Generalisierbarkeit der KI-Anwendungen auf die Gesamtbevölkerung (Murphy et al. 2021). In Bezug auf die Logopädie ist hier z.B. Repräsentativität von Kindersprachdaten im Hinblick auf den sozioökonomischen Hintergrund oder den Einbezug von Mehrsprachigkeit angesprochen.

Das Problem der Explainability (Erklärbarkeit) von Anwendungen Maschinellen Lernens, vor allem von Deep-Learning-Anwendungen bezieht sich auf deren zunehmend komplexen, mehrschichtigen Aufbau. Dieser führt dazu, dass es selbst für die Entwickler*innen nicht mehr nachvollziehbar ist, wie der Output dieser Systeme zustande kommt (Black-Box-Phänomen, Vellido 2019). Dies mag hinnehmbar sein, wenn es z.B. um personalisierte Werbung geht, die auf dem Bildschirm des eigenen Computers erscheint. Wenn allerdings Entscheidungen von der KI getroffen werden, die weitreichendere Konsequenzen haben, wie es bei einem Einsatz in der logopädischen Diagnostik der Fall wäre, ist eine solche Intransparenz nicht tragbar.

Die Einbindung von mehr oder weniger transparenter KI-Software in Entscheidungen birgt eine weitere ethisch-rechtliche Herausforderung. Wer ist verantwortlich – und somit auch haftbar – für mögliche Fehler? Auf diese Frage verweist der Aspekt der Accountability. Hier wird diskutiert, welche Ver-



Abrechnen, Termine planen und Patienten verwalten

Mit Severins und thevea geht das mühelos - perfekt für jede Logopädiepraxis!

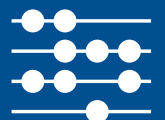


Severins GmbH

T. 0281 – 16394–50

meinangebot@severins.de

www.severins.de/logo



antwortung die Anwender*innen aus dem Gesundheitswesen aufgrund ihrer Zuständigkeit für Entscheidungen in Bezug auf die Gesundheit ihrer Patient*innen und welche Verantwortung Entwickler*innen für die sichere und fehlerfreie Anwendung von KI tragen (Murphy et al. 2021). Die Autor*innen der ethischen Richtlinien der WHO (2021) zu diesem Thema, „Ethics and governance of artificial intelligence for health“, tendieren klar dazu, Firmen, die KI-Software vermarkten, nicht aus der Verantwortung für ihre Produkte zu lassen. Die Frage nach der Haftbarkeit ist allerdings komplex und wird sich hoffentlich in Zukunft klären.

Datenschutz und Privatsphäre sind mit der zunehmenden Digitalisierung unseres täglichen Lebens in unser Bewusstsein gerückt. Wie bereits verdeutlicht, benötigten Anwendungen Maschinellen Lernens in der Entwicklung große Datenmengen und können als Endprodukt wiederum große Mengen an Daten verarbeiten (Murphy et al. 2021). Privacy ist also auch hier ein wichtiges Thema, z.B. wenn es um die Weitergabe oder Nutzung von elektronischen Patient*inrendaten durch kommerzielle Anbieter solcher Software in der Gesundheitsversorgung geht (Vellido 2019). In Abbildung 2 sind die beschriebenen ethisch-rechtlichen Herausforderungen zusammengefasst.

Perspektive für die Logopädie

Am Beispiel der logopädischen Diagnostik im Bereich Kindersprache wird die mögliche Arbeitserleichterung (z.B. durch die Automatisierung manueller Prozesse) und eine Erweiterung der diagnostischen Möglichkeiten (z.B. durch zusätzliche Auswertungsoptionen oder den Zugang zu anderen Erhebungskontexten) deutlich. Die allerdings ebenso

gegebenen Herausforderungen implizieren Handlungsbedarf.

- **Mündig werden:** Um sich eine Meinung zum zukünftigen Einsatz Künstlicher Intelligenz in der logopädischen Praxis bilden zu können, bedarf es Grundlagenwissen zur Funktionsweise, zu den Chancen und Risiken sowie zum Umgang mit dieser Technologie. Einführungen bieten z.B. die Beiträge von Ertel (2021), Taulli (2022) und Shen et al. 2021. Das Thema sollte in der logopädischen Aus- und Weiterbildung vermittelt werden.
- **Mitgestalten:** Der Einbezug von Software-Endnutzer*innen in der Entwicklung und Implementierung von KI wird in der Literatur vielfach gefordert (z.B. Shen et al. 2021). Praktiker*innen sind in der einzigartigen Position, Einsatzmöglichkeiten für KI zu identifizieren, die Modellentwicklung aus ihrer Perspektive zu leiten, Grenzen zu bestimmen, die Implementierung zu begleiten sowie die endgültige Nutzung der Software zu bewerten. Gleichzeitig sollten ethisch-rechtliche Richtlinien zukünftig je nach Anwendungsgebiet spezifiziert werden, um eine sichere, vertrauensvolle und effektive Nutzung von KI zu ermöglichen.
- **Rolle definieren:** Die Digitalisierung in der Logopädie wird durch den Einsatz Künstlicher Intelligenz auf ein völlig neues Niveau gehoben. War die Rolle der Software-Nutzer*innen in der Diagnostik bisher klar definiert, stellt die Entwicklung sich selbst optimierender Anwendungen Maschinellen Lernens, die in manchen Bereichen menschliche kognitive Leistungen übertreffen, unser Bild von Agierenden und Werkzeug – gefördert durch die Vermenschlichung dieser Technologie in den Medien – auf die Probe. Hier gilt es, die logopädische Expertise aus neuer Perspektive auszuloten, denn auch diese

Software bleibt ein Werkzeug. Den Wert menschlicher Beziehungen, Erfahrung und Wahrnehmung in Diagnostik und Therapie kann kein Computerprogramm ersetzen. Gleichwohl eröffnen sich uns durch einen progressiven Umgang mit KI Möglichkeiten, um die Versorgung unserer Patient*innen zu verbessern.

LITERATUR

- Alharbi, S., Alrazgan, M., Alrashed, A., Alnomasi, T., Almojel, R., Alharbi, R., ... & Almojil, M. (2021). Automatic Speech Recognition: Systematic Literature Review. *IEEE Access* 9, 131858–131876
- Chen, Y.-J. (2011). Identification of articulation error patterns using a novel dependence network. *IEEE Transactions on Biomedical Engineering* 58 (11), 3061–3068
- Ehlert, H., Beaulac, E., Gebauer, C., Rumberg, L., Ostermann, J. & Lüdtkke, U. (in Vorb.). *Towards automated assessment of speech sound disorders using connected speech: Exploring phonetic inventories*
- Ertel, W. (2021). Grundkurs Künstliche Intelligenz. Eine praxisorientierte Einführung. *Computational Intelligence*. Wiesbaden: Springer Vieweg
- Gale, R., Bird, J., Wang, Y., van Santen, J., Prud'hommeaux, E., Dolata, J. & Asgari, M. (2021). Automated scoring of tablet-administered expressive language tests. *Frontiers in Psychology* 12, 2986
- Gale, R., Dolata, J., Prud'hommeaux, E., van Santen, J., & Asgari, M. (2020). Automatic assessment of language ability in children with and without typical development. *42nd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*, 6111–6114
- Gilkerson, J. & Richards, J.A. (2008). The LENA™ developmental snapshot (Tech. Rep. No. LTR-07-2). Boulder, Connecticut: LENA Foundation
- Harrison, C.J. & Sidey-Gibbons, C.J. (2021). Machine learning in medicine: A practical introduction to natural language processing. *BMC Medical Research Methodology* 21 (1), 158
- Keshet, J. (2018). Automatic speech recognition: A primer for speech-language pathology researchers. *International Journal of Speech-Language Pathology* 20 (6), 599–609
- Klatte, I., van Heugten, V., Zwisserlood, R. & Gerrits, E. (2021). Language sample analysis in clinical practice: Speech-language pathologists' barriers, facilitators, and needs. *Language, Speech, and Hearing Services in Schools* 53, 1–16
- Lüdtkke, U., Bornman, J., de Wet, F., Heid, U., Ostermann, J., Rumberg, L., ... & Ehlert, H. (2023). Multidisciplinary perspectives on automatic analysis of children's language samples: Where do we go from here? *Folia Phoniatrica et Logopaedica* 75, 1–12

Abb. 2: Auswahl ethisch-rechtlicher Herausforderungen und Bedarfe beim Einsatz von KI im Gesundheitswesen

Fairness	Explainability	Accountability	Privacy
<p>Problem: Algorithmischer BIAS (Algorithmen spiegeln menschlichen BIAS oder erzeugen neuen), mangelhafte Trainingsdaten. Wichtig, z.B.: Repräsentativität der Trainingsdaten für den späteren Anwendungszweck</p>	<p>Problem: Mangelnde Interpretierbarkeit des Zustandekommens des KI-Outputs (black box) Wichtig, z.B.: Methoden, die den Anwendenden eine Bewertung des KI-Outputs ermöglichen</p>	<p>Problem: Haftung für mögliche Fehler der KI Wichtig, z.B.: Klare, bindende Richtlinien, Firmen, welche KI-Anwendungen vertreiben, nehmen ihre Verantwortung wahr</p>	<p>Problem: KI-Entwicklung und Anwendung nutzt große Mengen an Daten Wichtig, z.B.: Gleichgewicht zwischen Schutz der Privatsphäre und nachhaltiger Bereitstellung von Daten</p>

MacWhinney, B. (2000). The CHILDES project: Tools for analyzing talk. *Child Language Teaching and Therapy* 8 (2)

McKechnie, J., Shahin, M., Ahmed, B., McCabe, P., Arciuli, J. & Ballard, K. J. (2021). An automated lexical stress classification tool for assessing dysprosody in childhood apraxia of speech. *Brain Sciences* 11 (11), Article 11

McKechnie, J., Ahmed, B., Gutierrez-Osuna, R., Monroe, P., McCabe, P. & Ballard, K. J. (2018). Automated speech analysis tools for children's speech production: A systematic literature review. *International Journal of Speech-Language Pathology* 20 (6), 583-598

Miller, J.F. & Chapman, R. (1985). *Systematic analysis of language transcripts* [Computer software manual]. Madison, WI

Murphy, K., Di Ruggiero, E., Upshur, R., Willison, D.J., Malhotra, N., Cai, J.C., ... & Gibson, J. (2021). Artificial intelligence for good health: A scoping review of the ethics literature. *BMC Medical Ethics* 22 (1), 14

Ng, S.-I. & Lee, T. (2020). *Automatic detection of phonological errors in child speech using Siamese recurrent autoencoder*. arXiv:2008.03193

Ng, S.I., Tao, D., Wang, J., Jiang, Y., Ng, W.Y. & Lee, T. (2018). An automated assessment tool for child speech disorders. *11th International Symposium on Chinese Spoken Language Processing (ISCSLP)*, 493-494

Nadkarni, P.M., Ohno-Machado, L. & Chapman, W.W. (2011). Natural language processing: An introduction. *Journal of the American Medical Informatics Association* 18 (5), 544-551

Naziya, S. & Deshmukh, R.R. (2016). Speech recognition system – A review. *IOSR Journal of Computer Engineering* 18 (04), 1-9

Radford, A., Kim, J.W., Xu, T., Brockman, G., McLeavey, C. & Sutskever, I. (2022). *Robust Speech Recognition via Large-Scale Weak Supervision*. arXiv:2212.04356

Shahin, M., Ahmed, B., Parnandi, A., Karappa, V., McKechnie, J., Ballard, K.J. & Gutierrez-Osuna, R. (2015). Tabby Talks: An automated tool for the assessment of childhood apraxia of speech. *Speech Communication* 70, 49-64

Shen, L., Kann, B.H., Taylor, R.A. & Shung, D.L. (2021). The clinician's guide to the machine learning galaxy. *Frontiers in Physiology* 12, 658583

Shulga, D., Silber-Varod, V., Benson-Karai, D., Levi, O., Vashdi, E. & Lerner, A. (2020). Toward explainable automatic classification of children's speech disorders. In: Karpov, A. & Potapova, R. (Hrsg.), *Speech and Computer* (509-519). Cham: Springer International Publishing

Searle, J.R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences* 3 (3), 417-424

Spreer, M. (2018). *Diagnostik von Sprach- und Kommunikationsstörungen im Kindesalter*. München: Ernst Reinhardt Verlag

Taulli, T. (2022). *Grundlagen der Künstlichen Intelligenz: Eine nichttechnische Einführung*. Heidelberg: Springer

Topal, M.O., Bas, A. & van Heerden, M. (2021). *Exploring transformers in natural language generation: GPT, BERT, and XLNet*. arXiv:2102.08036

Vellido, A. (2019). Societal issues concerning the application of artificial intelligence in medicine. *Kidney Diseases* 5 (1), 11-17

WHO – World Health Organization (Hrsg.) (2021). *Ethics and governance of artificial intelligence for health: WHO Guidance*. Geneva: World Health Organization

Yu, K.-H., Beam, A.L. & Kohane, I.S. (2018). Artificial intelligence in healthcare. *Nature Biomedical Engineering* 2 (10), Article 10

Zandi, D., Reis, A., Vayena, E. & Goodman, K. (2019). New ethical challenges of digital technologies, machine learning and artificial intelligence in public health: A call for papers. *Bulletin of the World Health Organization* 97 (1), 2-2

Dr. Hanna Ehlert

Logopädin und wissenschaftliche Mitarbeiterin an der Leibniz Universität Hannover. Sie habilitiert im interdisziplinären TALC-Projekt, in welchem der Einsatz und die ethischen Rahmenbedingungen von Anwendungen Maschinellen Lernens in der Logopädie erforscht werden.



Prof. Dr. Ulrike Lütcke

Professorin an der Leibniz Universität Hannover. Dort leitet sie die Studiengänge für Sprachtherapie und Sonderpädagogik (Schwerpunkt Sprache), sowie die Hochschulambulanz und die Forschungsinitiative „Leibniz Lab for Relational Communication Research“. Ihr Forschungsschwerpunkt liegt in der Bedeutung von Emotionen für den Spracherwerb. Weiterhin engagiert sie sich im Aufbau der Profession Logopädie/Sprachtherapie in Subsahara Afrika.



SUMMARY. Artificial intelligence in speech and language therapy. Potential and challenges of automation in pediatric speech and language assessment

The article addresses AI-based automation in the assessment of child speech and language. AI, machine Learning and the two digital applications relevant for this purpose are introduced: automatic speech recognition (ASR) and automatic text processing. Following this, advances and limitations of this technology are presented with regard to automated speech assessment, automated test score interpretation and automated language sample analysis. Additionally, technical and ethical challenges of automation for clinical assessment purposes and. Finally, needs for a productive and meaningful use of this technology in speech and language therapy for children are pointed out.

KEYWORDS: artificial intelligence (AI) – machine learning – pediatric speech and language assessment – automatic speech recognition – natural language processing

<https://doi.org/10.2443/skv-s-2023-53020230601>

KONTAKT

Dr. phil. Hanna Ehlert

Leibniz Universität Hannover
Schlosswender Straße 1
30159 Hannover
hanna.ehlert@ifs.uni-hannover.de

Anzeige



Schluckdiagnostiksysteme speziell für LogopädInnen und SprachtherapeutInnen vom Experten

<https://www.rehder.de/logo>

