

Martin Steppan, Marc Birkhölzer

High resolution psychotherapy research durch KI-gestützte Prozessanalyse



Psychotherapie-Wissenschaft
15. Jahrgang, Nr. 1, 2025, Seite 29–38
DOI: 10.30820/1664-9583-2025-1-29
Psychosozial-Verlag

Impressum

Psychotherapie-Wissenschaft
ISSN 1664-9591 (digitale Version)
15. Jahrgang Heft 1/2025
<https://doi.org/10.30820/1664-9583-2025-1>
info@psychotherapie-wissenschaft.info
www.psychotherapie-wissenschaft.info

Herausgeber

Assoziation Schweizer Psychotherapeutinnen
und Psychotherapeuten
Geschäftsstelle ASP
Riedtlistr. 8
CH-8006 Zürich
Tel. +41 43 268 93 00
www.psychotherapie.ch

Redaktion

Mara Foppoli, Lugano
Lea-Sophie Richter, Zürich
Mario Schlegel, Zürich
Peter Schulthess, Zürich

Hinweise für AutorInnen befinden
sich auf der Homepage der Zeitschrift:
www.psychotherapie-wissenschaft.info

Verlag

Psychosozial-Verlag GmbH & Co. KG
Walltorstr. 10
D-35390 Gießen
+49 6421 96 99 78 26
info@psychosozial-verlag.de
www.psychosozial-verlag.de

Anzeigen

Anfragen zu Anzeigen bitte an den Verlag:
anzeigen@psychosozial-verlag.de
Es gelten die Preise der auf www.psychosozial-verlag.de
einschbaren Mediadaten.
ASP-Mitglieder wenden sich bitte direkt an
die ASP-Geschäftsstelle: asp@psychotherapie.ch

Datenbanken

Die Zeitschrift Psychotherapie-Wissenschaft wird regelmäßig
in der Abstract-Datenbank PsycInfo der American Psycho-
logical Association (APA), im Directory of Open Journals
(DOAJ) und in den Publikationsdatenbanken PSYINDEX und
PsychArchives des Leibniz-Institut für Psychologie/Leibniz
Institute for Psychology (ZPID) erfasst.



Die Beiträge dieser Zeitschrift sind unter der Creative Commons
Attribution-NonCommercial-NoDerivs 3.0 DE Lizenz lizenziert.
Diese Lizenz erlaubt die private Nutzung und unveränderte
Weitergabe, verbietet jedoch die Bearbeitung und kommerzielle
Nutzung. Weitere Informationen finden Sie unter:
creativecommons.org/licenses/by-nc-nd/3.0/de

High resolution psychotherapy research durch KI-gestützte Prozessanalyse

Ein Paradigmenwechsel in der Psychotherapieforschung

Martin Steppan & Marc Birkhölzer

Psychotherapie-Wissenschaft 15 (1) 2025 29–38

www.psychotherapie-wissenschaft.info

CC BY-NC-ND

<https://doi.org/10.30820/1664-9583-2025-1-29>

Zusammenfassung: Die Psychotherapieforschung erlebt durch den Einsatz Künstlicher Intelligenz (KI) einen grundlegenden Paradigmenwechsel. Während Paul Ekman's Arbeiten zur Emotionserkennung auf der Beobachtung und manuellen Kodierung von Mimik basierten, ermöglicht KI heute eine automatisierte, multimodale Analyse von Sprache, Mimik, Gestik und emotionaler Dynamik. Diese Technologien erlauben erstmals eine beinahe vollständige Erfassung, Dokumentation und Analyse gesamter Therapiesitzungen in Echtzeit. Der Beitrag untersucht den Übergang von Ekman's methodischem Ansatz zur KI-gestützten «Totalerhebung» und analysiert die wissenschaftlichen sowie klinischen Implikationen. Besonderes Augenmerk wird auf die Chancen gelegt, die sich aus der hochauflösenden Echtzeitanalyse therapeutischer Prozesse ergeben, sowie auf die Risiken im Hinblick auf Datenschutz und ethische Standards. Abschliessend wird diskutiert, wie durch Anonymisierung und strenge Sicherheitsmassnahmen ein verantwortungsvoller Umgang mit sensiblen Therapiedaten gewährleistet werden kann, um die Potenziale dieser Technologien voll auszuschöpfen.

Schlüsselwörter: Künstliche Intelligenz, Erkenntnistheorie, Emotionserkennung, affektive Datenverarbeitung, klinische Überwachung, Psychotherapie, Psychiatrie, neue Methoden

Die Psychotherapieforschung steht an der Schwelle eines fundamentalen Wandels bzw. Paradigmenwechsels: Während früher die Erfassung von Körpersprache und Emotionen auf Beobachtungen und manuellen Analysen basierte, ermöglicht KI heute eine automatisierte Analyse von Sprache, Mimik und Gestik, wobei dies sogar in Echtzeit möglich ist. Diese Technologien erlauben eine beinahe vollständige Erfassung der verbalen und non-verbalen Kommunikation zwischen Therapeut und Patient während der gesamten Therapiesitzung.

Ein wichtiger Schritt in der Geschichte der Körpersprache-Forschung war Paul Ekman's Facial Action Coding System (FACS), das die Analyse der Gesichtsmimik standardisierte (Ekman & Friesen, 1974). Doch diese Methode war auf kurze Videoabschnitte und aufwendige manuelle Kodierungen beschränkt. Die Psychotherapieforschung mithilfe dieser Methode war daher eine sehr zeitaufwändige Angelegenheit, die nur eine fragmentarische und selektive Erfassung von Körpersprache erlaubte (Merten et al., 1996). Dank KI kann heute eine beinahe lückenlose Erfassung der Körpersprache in Echtzeit stattfinden (Jaiswal & Nandi, 2020). Dabei können nicht nur Mimik, sondern auch Gestik, Körperhaltung, Fingerbewegungen oder Blickbewegungen simultan analysiert werden (Singh et al., 2022). In Kombination mit der ebenfalls akkuraten automatisierten Spracherkennung (Arik et al., 2017) bieten sich hier völlig neue Möglich-

keiten, die emotionalen Prozesse zwischen Patient und Therapeutin zu untersuchen.

Dass es sich hierbei nicht nur um eine inkrementelle Steigerung der verfügbaren Daten handelt, sondern einen «Quantensprung» in den Möglichkeiten, d. h. um einen möglichen «Paradigmenwechsel», ist Gegenstand dieses Artikels. Mit diesen neuen Möglichkeiten stellt die Nutzung dieser Technologien die Forschung auch vor neue Herausforderungen, insbesondere im Bereich Datenanalyse, Datenschutz und Ethik. Wie kann sichergestellt werden, dass sensible Therapiedaten sicher und verantwortungsvoll genutzt werden? Ziel dieses Beitrags ist es, die wissenschaftlichen Potenziale der KI-gestützten Analysen zu zeigen und gleichzeitig die methodischen und ethischen Risiken kritisch zu beleuchten.

Forschung zu Körpersprache in der Psychotherapie vor KI

Die wissenschaftliche Erforschung der Mimik und Körpersprache in der Psychotherapie hat eine lange Tradition (Darwin, 1872). Während zunächst vor allem Methoden der Beobachtung und Beschreibung des menschlichen nonverbalen Verhaltens im Vordergrund standen (Mehrabian & Ferris, 1967), rückten mit der Zeit immer mehr Messverfahren in der Vordergrund. Massgeblich

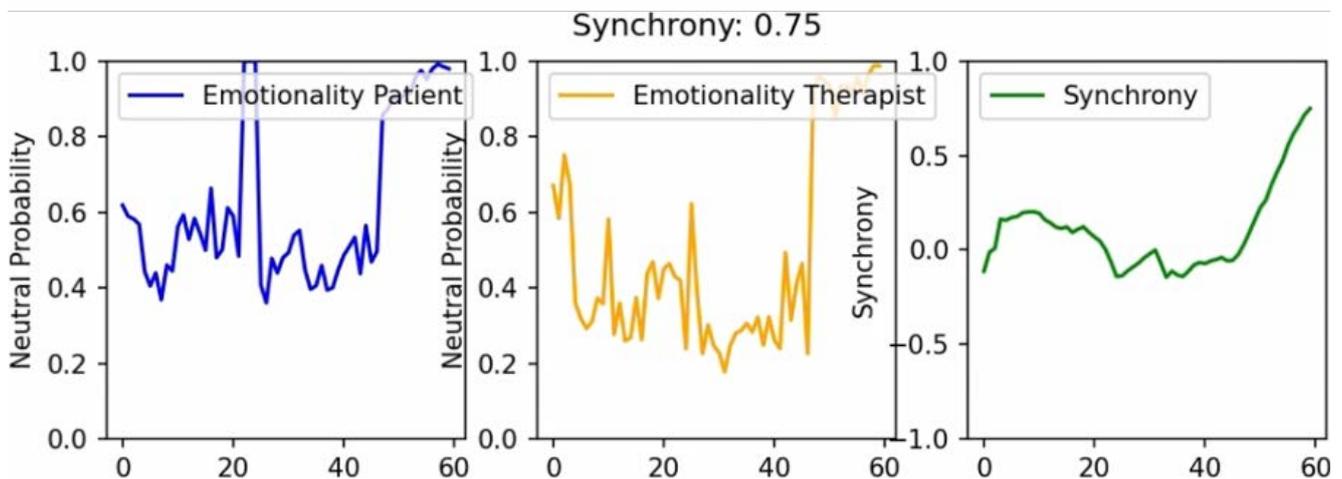
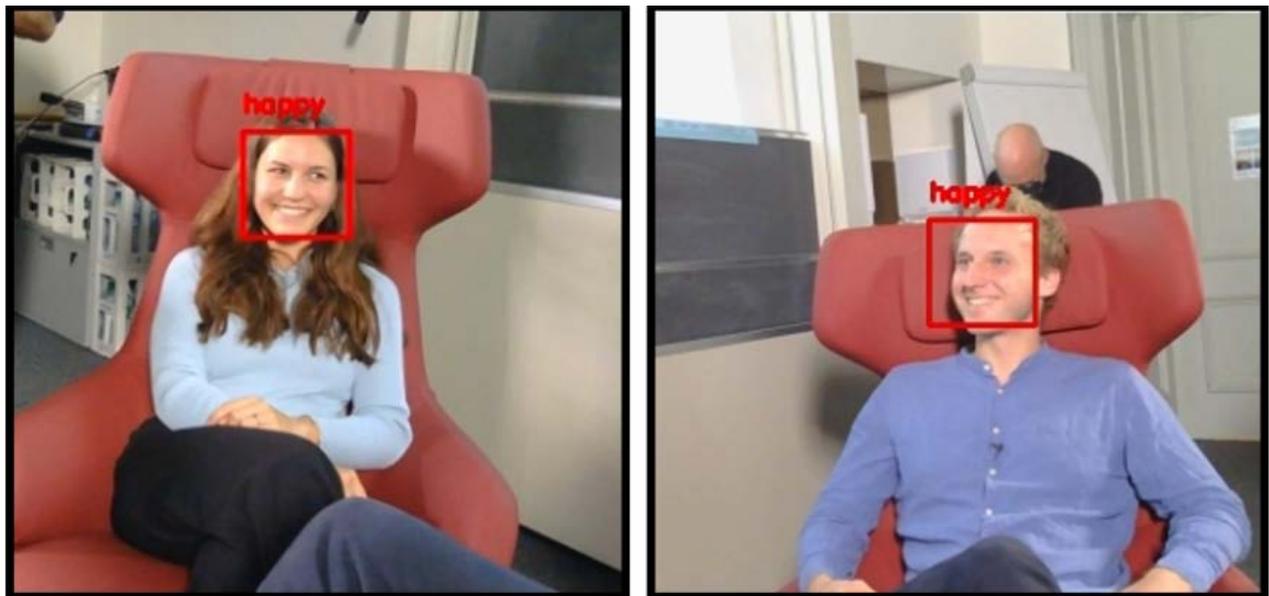


Abb. 1: Wenig invasives Setup des Psychotherapielabors im Rahmen der AIT-Studie an den Universitären Psychiatrischen Kliniken Basel, registriert bei clinicaltrials.gov (NCT02518906). Automatische Emotionserkennung und Berechnung des Spiegels von Emotionen ist in Echtzeit möglich (Nachstellung der Therapiesituation für Dreharbeiten an den Universitären Psychiatrischen Kliniken Basel).

verantwortlich dafür ist Paul Ekman mit dem Facial Action Coding System (FACS), das eine standardisierte Kategorisierung der Gesichtsmimik gestattet (Ekman & Friesen, 1978).

Die Kodierung einzelner Standbilder oder Videosequenzen mit dem FACS ist mit erheblichem Arbeits- und Zeitaufwand verbunden: Je nach Auflösung und emotionaler Aktivität, wurde ein etwa 100-facher Zeitaufwand berichtet, also ca. 100 Minuten für ein einminütiges Video (Cohn et al., 2007). Die ersten Studien beschäftigten sich daher eher mit der Kodierung der Mimik in der Monade, d.h. nur einer einzelnen Person. Wichtige Studien aus dieser Zeit beschäftigten sich bspw. mit neurobiologischen Korrelaten der Gesichtsausdrücke (Rinn, 1984) oder der hohen kulturellen Universalität von Basisemotionen (Ekman & Friesen, 1971). Umso zeitaufwändiger waren naturgemäß die ersten Anwendungen dieses Ansatzes in der dyadischen Forschung. Besonders relevant waren hier

zunächst die Erforschung der nonverbalen Interaktion zwischen Mutter und Kind als Marker für die Bindungsentwicklung (Beebe, 1982). Diese Studien führten in weiterer Folge auch zum *still face paradigm* – also zu jenen bekannten Experimenten, in denen das Ausbleiben einer mimischen Reaktion zu starken Reaktionen der Säuglinge führte (Tronick et al., 1978). Diese Befunde belegten, wie wichtig die nonverbale Kommunikation und emotionale Resonanz bereits im frühkindlichen Alter ist.

In weiterer Folge entwickelte sich seit den 1980ern auch psychotherapeutische Forschung, die u. a. auch besonders aus psychoanalytischer Richtung motiviert war, gerade auch weil in der Körpersprache und in der Mimik unbewusste Signale, z.B. durch Mikroexpressionen, zutage treten können. Federführend in dieser Hinsicht waren im deutschen Sprachraum Eva Bänninger-Huber und Kollegen (Bänninger-Huber & Steiner, 1992). Trotz des erheblichen Aufwands, der mit diesen Studien ver-

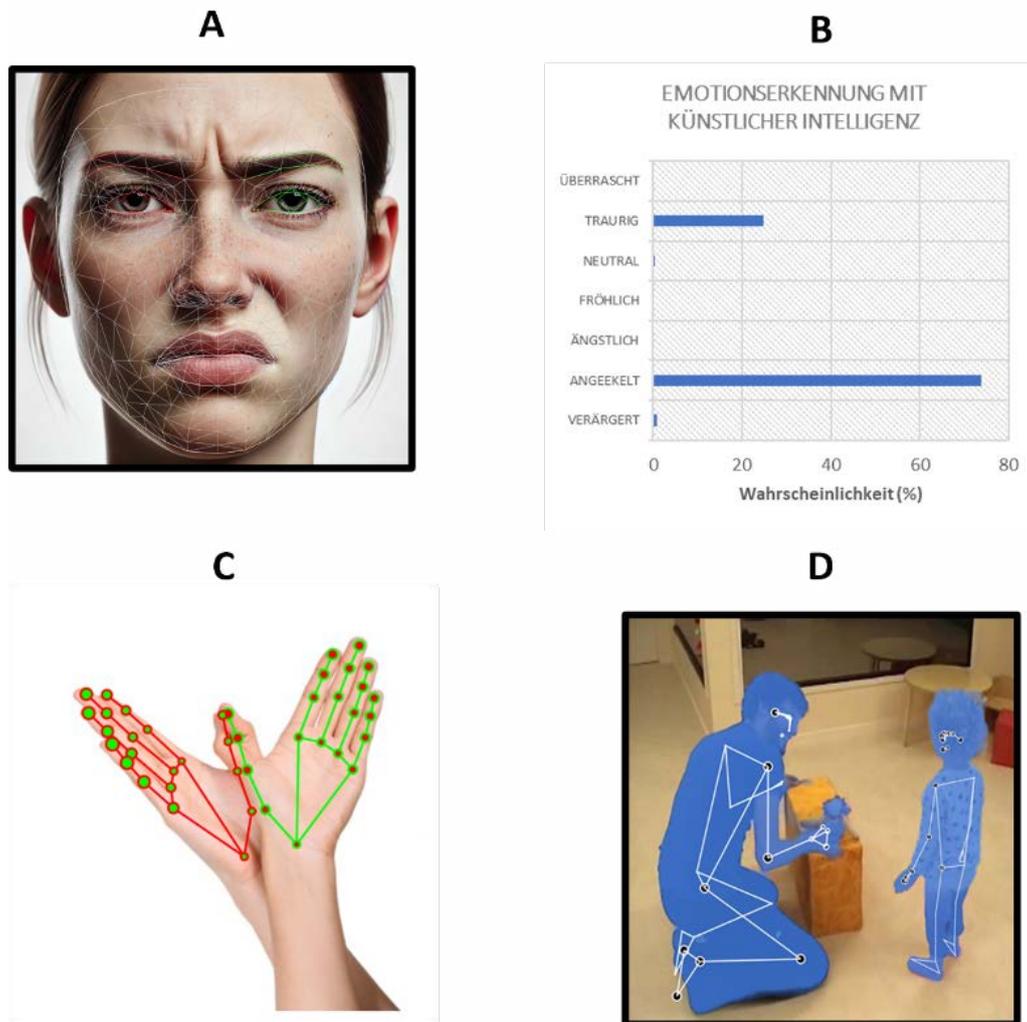


Abb. 2: Beispiele der umfangreichen Analyse der Körpersprache durch KI in Echtzeit: (A) automatische Erkennung von landmark points auf dem Gesicht und Avatarisierung von Gesichtern zur Sicherung des Datenschutzes; (B) automatische Emotionserkennung auf dem Gesicht links; (C) automatische Erkennung von Hand- und Fingerbewegungen; (D) automatische Erkennung von wichtigen Körperpunkten eines Therapeuten mit einem Kind im Spiel (mit freundlicher Genehmigung des Therapiezentrums FIAS in Muttenz, Schweiz)

bunden war, konnten hier entscheidende «mikroanalytische» Konzepte entwickelt werden: Mit «prototypischen affektiven Mikrosequenzen» sind kurze unbewusste, emotionale Abläufe gemeint, die dazu dienen, Störungen in der Affektregulation in Beziehungen auszubalancieren. Solche Mikrosequenzen sind häufig durch ein Lächeln oder andere Formen der zugewandten Körpersprache gekennzeichnet, die durch positive Resonanz die affektive Bindung zweier Interaktionspartner erhöhen. Dabei stehen Therapeuten gemäss Bänninger-Huber und Kollegen vor dem Balanceakt, einerseits die Beziehung zu stabilisieren und andererseits eine notwendige «Konfliktspannung» aufrechtzuerhalten (Bänninger-Huber, 2021). Nicht weniger interessant ist das in diesem Zusammenhang entwickelte Konzept der *trap*, d. h. einer «Falle». Diese zeigt sich nicht nur in verbalen, sondern auch in nonverbalen Verhaltensmustern, wie etwa durch Körpersprache, Mimik oder Gestik. Diese nonverbalen Signale

werden als «Falle» bezeichnet, da sie das Rollenangebot der Patientin unterstützen, indem sie die Erwartungen und emotionalen Bedürfnisse subtil kommunizieren und Therapeutinnen dazu verleiten sollen, auf eine bestimmte Weise zu reagieren (Bänninger-Huber & Widmer, 2020). Besonders aufschlussreich sind in diesem Zusammenhang auch jene Studien, die das mimisch-affektive Verhalten in psychologisch relevanten Situationen, z. B. während des Adult Attachment Interviews untersuchen (Buchheim & Benecke, 2007). Einige dieser Studien verwendeten hier bereits das EMFACS (also Emotional Facial Action Coding System), wobei auch hier bereits erste Methoden der Automatisierung zum Einsatz kamen (Benecke & Krause, 2005).

Abgesehen von der Gesichtsmimik gab es immer wieder auch Versuche, allgemeine körpersprachliche Signale wie Körperhaltung, Gestik oder dergleichen standardisiert zu erfassen, z. B. mithilfe des Body Action and Posture Co-

ding System (Dael et al., 2012). Manuell betrieben ist die Forschung hier ähnlich aufwendig, und hat zusätzlich das Problem, dass die Kodierungsschemata weniger etabliert sind als bspw. das FACS. Umso mehr hervorzuheben sind neuere Ansätze, die zwar noch nicht KI im strengen Sinne benutzen, aber mithilfe z. B. von Bewegungssensoren körperliche Bewegungsmuster von Patienten und Therapeutinnen erfassen. Besonders methodisch interessant ist hierbei die Motion Energy Analysis (MEA) von Fabian Ramseyer (2020), die dies ohne Bewegungssensoren anhand von Videodaten ermittelt. Von einer exakten Messung der Bewegungen, z. B. der Gliedmassen oder Finger, sind diese Ansätze zwar noch ein Stück entfernt, nichtsdestotrotz konnte man diese Daten bereits heranziehen, um z. B. die Bewegungssynchronizität zu quantifizieren, die positiv mit dem Therapieerfolg assoziiert ist (Zimmermann et al., 2021).

Die methodischen Grenzen dieser Ansätze, die noch ohne KI oder andere maschinelle Automatisierung (mit Ausnahme von Motion Energy Analysis) auskommen, sind dabei schnell ersichtlich: (a) mangelnde Ökonomie und zum Teil grosser menschlicher Aufwand; (b) damit verbundene hohe Kosten und anstrengende Tätigkeit für menschliche Evaluation; (c) eingeschränkte Reproduzierbarkeit im Sinne von Beobachterabhängigkeit und nicht perfekter Inter-Rater-Übereinstimmung. (d) Aufgrund des hohen Aufwandes konnten diese Formen der Messung von Körpersprache zum Teil nur auf Sequenzen oder Stichproben aus Videos durchgeführt werden.

Die Revolution durch KI: Automatisierte Emotionserkennung oder *affective computing*

Eine der zentralen Domänen von KI ist Mustererkennung. Die menschliche Mimik, die als Interaktionsform stammesgeschichtlich älter ist als die verbale Kommunikation (De Waal, 2014), erfüllt exakt diese Anforderung eines Musters, das durch simulierte Nervensysteme (neuronale Netze) erkannt werden kann. Dabei sind mittlerweile die einfachsten solcher Netze bereits besser in der Klassifikation als menschliche Rater (Steppan et al., 2024). Die dafür benötigte Rechenleistung ist dementsprechend gering, und es können mit einem herkömmlichen Computer die Bilder sowohl von Patientinnen als auch Therapeutinnen in einer Auflösung von 30 Bildern/Sekunde ohne Weiteres in Echtzeit analysiert werden – somit ein «Quantensprung» seit Ekmans Methoden (Abb. 1 zeigt einen Screenshot der Nachstellung einer solchen Situation).

Ein weiterer selten beleuchteter Bereich der automatisierten Analyse der Körpersprache umfasst die automatisierte Gliedmassenentdeckung (*limb detection*, s. Abb. 2C & 2D). Es ist also möglich, in derselben Geschwindigkeit Schultern, Arme, Beine, Kopf und Finger mithilfe von KI zu lokalisieren. Es ist somit ein beinahe vollständiges 3D-Abbild der körperlichen Bewegungen und Gesten möglich. Dies inkludiert Verhaltensweisen, die in der Psychiatrie und Psychotherapie relevant sein könnten, z. B. repetitive Bewegungsmuster, Tics etc.

Darüber hinaus zählen zum *affective computing* auch all jene Ansätze, die versuchen, basierend auf dem rein akustischen Signal der Stimme, das emotionale Arousal zu bestimmen, z. B. durch den Pitch (also die Tonhöhe) oder durch Schwankungen der Tonhöhe, um nur einige Marker zu nennen. Zentrale Bedeutung hat hier naturgemäss die Trennung der Signale der Stimme von Patient und Therapeut (Fürer et al., 2020).

Ausgeklammert seien hier die enormen Fortschritte im Bereich der automatisierten Spracherkennung, die bereits jetzt eine vollständige Automatisierung der Transkription von Psychotherapien erlauben (Miner et al., 2020). Die Entwicklungen gehen bereits soweit, dass auch Mundarten oder Varietäten (z. B. Schweizerdeutsch) diesen Algorithmen keine Probleme bereiten (Arabsky et al., 2021). Dementsprechend gross sind die Anwendungsmöglichkeiten, wenn noch komplexere Modelle zur Textanalyse (z. B. *large language models*) angewandt werden. Erste Anbieter sind bereits auf dem Markt aktiv, die eine automatische Transkription erlauben, und darüber hinaus es ermöglichen, den gesamten Therapieverlauf nach Themenkomplexen (z. B. «Angst», «Medikation», spezifische Personen etc.) zu durchsuchen (z. B. <https://eleos.health>). Dementsprechend vielversprechend sind auch jene Innovationen, die zeigen, dass das Gesagte auch automatisiert auf kognitive Verzerrungen, ethisch problematisches Verhalten von Therapeutinnen etc. untersucht werden kann (Chen et al., 2023). Im gleichen Ausmass wie gesprochene Sprache komplexer ist als Körpersprache, können in diesem Bereich womöglich noch tiefgreifendere Ergebnisse erwartet werden.

Von der Stichprobe zur Totalerhebung: Big Data in der Psychotherapieforschung

Man stelle sich nunmehr vor, dass heute mit relativ einfachen Mitteln, d. h. zwei Webcams, Mikrofonen) ein beinahe vollständiges Abbild der Körpersprache (sowie des Gesagten) in der Form möglich ist, dass man die Szene in einen Avatar oder Trickfilm verwandeln könnte. Darüber hinaus ist auch alles Gesagte in der Form zugänglich, dass sämtliches Gesprochene automatisch transkribiert werden kann, und somit durchsuchbar, abstrahierbar und semantisch analysierbar ist.

Es ist davon auszugehen, dass dieser neue Reichtum an Daten eine nicht bloss quantitative Steigerung der Daten darstellt. Während vor einigen Jahren vielleicht eine kurze Videosequenz stichprobenartig in ähnlicher Weise ausgewertet werden konnte, lässt sich dies heute für ganze Psychotherapien und ganze Archive von Videomaterial vollständig durchführen. Vergleichbar ist diese Steigerung des Datenvolumens mit dem Übergang vom *candidate gene approach* in der Genetik, hinzu den genomweiten Assoziationsstudien (Duncan et al., 2019). Während noch vor wenigen Jahren bei einzelnen vielversprechenden genetischen Markern gesucht wurde, kann heute ohne vorhergehende Hypothese das ganze relevante Erbgut durchsucht werden. Eine vergleichbare methodische Herausforderung stellt sich mit den neuen

Psychotherapiedaten auch. Wie in der Genetik geht diese Form der Vollerhebung mit einer anderen Form der Datenauswahl, Hypothesenbildung und Statistik einher. Die Forschung in diesem Bereich steht somit vor einem grundlegenden methodischen Wandel.

Methodologische Implikationen: Der «Fluch der Dimensionalität»

Mit der neuen Methodologie ergeben sich auch neue methodische Rahmenbedingungen. Ein wesentlicher Vorteil der KI-gestützten Emotionserkennung ist ihre Reproduzierbarkeit. Sofern nicht anders eingestellt, sind die Algorithmen deterministisch, d. h. der gleiche Input kommt stets zum selben Output, und damit sind die Ergebnisse jederzeit überprüfbar und replizierbar (Bajwa et al., 2020). Dies steht der deutlich geringeren Übereinstimmung von menschlichen Einschätzungen gegenüber (Cabitza et al., 2022). D. h., dass in der Theorie die modernen Algorithmen vollständig reliabel (im Sinne von reproduzierbar) sind. Die KI-Methoden erfüllen somit strenge wissenschaftliche Kriterien hinsichtlich ihrer grundsätzlichen Überprüfbarkeit. Auf der anderen Seite werden diese Arten von Algorithmen trotzdem oft als *black boxes* bezeichnet, d. h., ein genaues Verständnis, wie der Algorithmus zu einem Ergebnis kommt, ist nur sehr schwer möglich und somit oft intransparent (Agrawal & Mittal, 2020). Darüber hinaus sind die Algorithmen oft in der Lage, menschliche Rater zu übertrumpfen was Emotionserkennung anlangt (Eskimez et al., 2016). In anderen Worten: Die KI-gestützten Methoden haben potenziell höchste Reliabilität und Reproduzierbarkeit, z. T. höhere externe Validität als die menschliche Einschätzung, sind aber nicht im letzten Detail nachvollziehbar. Diese Einschränkung trifft freilich auch für Menschen zu, da auch das menschliche Urteil nicht in die letzte Tiefe ergründbar ist. Demgegenüber sind Algorithmen zumindest potenziell transparent, d. h., der Code kann im Detail analysiert werden.

Aufgrund der grossen Datenfülle durch KI (und andere Biosensoren) stellen sich nun ähnliche Probleme, wie sie auch aus anderen Forschungsbereichen bekannt sind, und oft als «Fluch der Dimensionalität» bezeichnet werden (Chattopadhyay & Lu, 2019). Vereinfacht ausgedrückt liegt ein solcher vor, wenn die Anzahl der Personen deutlich geringer ist als die Anzahl der gemessenen Variablen (Altman & Krzywinski, 2018). Dieses Problem lag bspw. in den Anfangszeiten der genomweiten Assoziationsstudien (GWAS) vor, wo man sehr viele Marker identifizieren konnte, aber erst wenige Personen untersucht hatte. Eine ähnliche Problematik liegt auch in den Neurowissenschaften schon seit Jahren vor, wo man mit hoch-dimensionierten Daten zu tun hat, die oft von wesentlich weniger Personen stammen (Berisha et al., 2021).

Dieser «Fluch des Big Data» begegnet einem auch in der Psychotherapieforschung: Besonders die Daten von Biosensoren (z. B. EKG, EEG, Hautleitfähigkeit etc.) sowie jene der beschriebenen KI-Algorithmen wie Emotionserkennung in Stimme und Mimik liefern schier unend-

liche Datenmengen, die jedoch in nur wenigen Personen «genestet» sind (McNeish, 2014). Die Herausforderung wird durch den hohen zeitlichen und personellen Aufwand der Datenerhebung in der Psychotherapie noch verstärkt. Anders als in biologischen Studien, in denen eine Speichelprobe genügt, erfordert die Sammlung psychotherapeutischer Daten aktive Beteiligung von Therapeutinnen und Patientinnen – ein limitierender Faktor.

Der *curse of dimensionality* äussert sich in mehreren methodischen Problembereichen: (A) *Datenkomplexität und Signalqualität*: Hochaufgelöste und heterogene Datenmengen sind oft redundant, verrauscht oder schwer zu integrieren, was die Identifikation relevanter Muster erschwert. Man spricht hier auch von *sparseness*, also Ausdünnung der Daten (ebd.). (B) *Modellierung hochdimensionaler Prozesse*: Die Vielzahl an Variablen erhöht das Risiko von *overfitting* (d. h. Überanpassung der Modelle an Daten) und numerischer Instabilität (Orrù et al., 2020). Effiziente Reduktion oder Auswahl der Daten ist notwendig, um komplexe Beziehungen modellierbar zu machen, ohne wesentliche Informationen zu verlieren. (C) *Praxisnähe und Interpretierbarkeit*: Black-Box-Modelle sind schwer nachvollziehbar und begrenzen die Anwendbarkeit in der Praxis. Ergebnisse müssen verständlich und generalisierbar auf Patienten zugeschnitten sein, um therapeutisch nutzbar zu sein.

Ein Praxisbeispiel und methodische Herausforderungen

Stellen wir uns vor, eine Forscherin analysiert ein komplettes Archiv an Psychotherapievideos mithilfe KI-gestützter Tools. Dabei könnten sowohl die verbale Kommunikation als auch subtile Aspekte der Körpersprache sowie eine Reihe von hochauflösenden Biomarkern (z. B. EEG, EKG) automatisiert ausgewertet werden. Theoretisch liessen sich so beinahe unendliche Interaktionsformen und Datenkombinationen untersuchen und deren Zusammenhänge mit Therapieerfolgen explorieren. Doch genau hier liegt die Gefahr: Die schiere Fülle an Daten lädt dazu ein, beliebige Hypothesen zu testen, bis ein signifikanter Zusammenhang gefunden wird – ein Phänomen, das als *p-hacking* bekannt ist (Reis & Fries, 2022). In der Genetik führte dieses Problem zur Notwendigkeit sehr grosser Biobanken mit mehreren hundert Tausenden Teilnehmern (z. B. UK Biobank). Die grosse Anzahl verfügbarer Signale führte auch zur Anpassung des Signifikanzlevels auf $p < 5 \times 10^{-8}$ – also eine Million Mal strenger als $p < .05$ –, um für multiples Testen zu kontrollieren (Tyrrell et al., 2021). Die Psychotherapieforschung steht damit vor einer doppelten Herausforderung: Einerseits müssen strenge methodische Standards – wie die Korrektur für multiples Testen – eingehalten werden. Andererseits besteht in diesem Bereich nicht im selben Ausmass die Möglichkeit, die Stichproben so deutlich zu erhöhen, wie dies in anderen Wissenschaften der Fall ist. Welcher Weg hier eingeschlagen werden kann, ist eine wesentliche Frage, die Psychotherapieforscher beschäftigen sollte.

Chancen durch KI-gestützte Analysen

Es ist unbestritten, dass Methoden der KI nicht nur ein theoretisches Potenzial in der medizinischen Forschung besitzen, sondern bereits heute in der Praxis beeindruckende Ergebnisse liefern. Besonders in der Radiologie und neurologischen Bildgebung – etwa bei der Segmentierung von Gehirntumoren – ist der klinische Nutzen längst evident (Najjar, 2023; Ranjbarzadeh et al., 2023). Diese Erfolge beruhen auf der Fähigkeit von KI-Systemen, optische Signale präzise zu analysieren. Es ist daher wenig verwunderlich, dass Bilddaten in diesen Disziplinen den zentralen Fokus bilden.

Auch in der Psychotherapie könnten KI-gestützte Methoden eine Schlüsselrolle einnehmen. Denn obwohl es hier weniger um statische Bilddaten als vielmehr um komplexe, dynamische Prozesse wie Körpersprache, Mimik und Sprache geht, ist gerade diese Vielschichtigkeit eine Stärke moderner Algorithmen. Die Analyse solcher Daten ist für Menschen äusserst aufwendig und zeitintensiv, aber für KI-Systeme geradezu prädestiniert. Dies eröffnet ein breites Spektrum neuer Möglichkeiten:

1. *Effizienz und Skalierbarkeit:* Wesentliche Probleme in der Versorgung von psychischen Problemen sind die mangelnde Zugangsgerechtigkeit sowie Fragen der Unter- oder Fehlversorgung (Fiske et al., 2019). Aufgrund seiner einfachen Skalierbarkeit für eine grosse Anzahl von Personen (z.B. durch Handy oder Webapps) birgt die KI hier naturgemäss grosses Potenzial. Zudem können KI-Lösungen besonders geeignet sein bei tabuisierten Themen (z.B. Paraphilien) oder auch bei *hard-to-reach populations*, also bei schwer zu erreichenden Zielgruppen. Dazu zählen z.B. Regionen, in denen die medizinisch-psychologische Grundversorgung schlecht ist. KI-Apps könnten hier in der Lage sein den «Hürdenlauf» zur Psychotherapie (Darboven et al., 2024) etwas weniger abschreckend zu gestalten oder auch Personengruppen zu motivieren, die sonst seltener Hilfe in Anspruch, z. B. Männer (Riffer et al., 2021).

2. *Neue Einsichten in dyadische Prozesse (Patient-Therapeutin):* Die Kombination von Körpersprache, Mimik und verbaler Sprache liefert tiefergehende Einblicke in die Interaktion zwischen Patientin und Therapeut. KI kann unbewusste Prozesse sichtbar machen, etwa Diskrepanzen zwischen tatsächlich gefühltem und verbal ausgedrücktem Zustand, oder die Authentizität beider Beteiligten analysieren (Park et al., 2021; Saxen et al., 2017). Solche Erkenntnisse könnten helfen, therapeutische Dissoziationen zu verringern – also die Kluft zwischen dem, was eine Person fühlt, und dem, was sie sagt. Eine grössere Übereinstimmung dieser Ebenen könnte nicht nur die Authentizität steigern, sondern auch als Indikator für Therapieerfolg dienen.

3. *Echtzeit-Rückmeldungen in der Therapie:* KI-gestützte Systeme könnten während der Sitzung unbewusste Signale oder emotionale Zustände analysieren und direktes Feedback liefern – sei es zur Wahrnehmung von Fortschritten oder zur Identifikation von Blockaden. Solche Echtzeit-Analysen könnten nicht nur die therapeutische Arbeit optimieren, sondern auch Patientinnen helfen, sich

selbst bewusster wahrzunehmen. Erste Echtzeit-Methoden sind bereits in Erprobung (Lin et al., 2022).

4. *Langfristige Beobachtung emotionaler Veränderungen:* Über mehrere Sitzungen hinweg kann KI emotionale und kommunikative Veränderungen dokumentieren und analysieren. Langfristige Muster und subtile Veränderungen, die für Menschen oft schwer erkennbar sind, könnten systematisch aufgezeigt werden. Diese Art von «Langzeit-Gedächtnis» der KI könnte besonders hilfreich sein, um den Therapieerfolg über Zeiträume hinweg zu evaluieren. Erste solche Datenquellen werden bereits gesammelt (Steppan et al., 2024; Tal et al., 2024).

5. *Objektivere Analysen durch Unvoreingenommenheit der KI:* Anders als Therapeutinnen ist KI unbeeinflusst von Gegenübertragungen oder emotionalen Reaktionen, die durch die Dynamik mit dem Patienten ausgelöst werden könnten. KI wird nicht getriggert und bleibt frei von persönlichen Erfahrungen oder Vorurteilen. Auch wenn die Technologie nicht völlig fehlerfrei ist (Han et al., 2020), so kann man doch von einer objektiveren und damit anwenderunabhängigen Qualität ausgehen im Vergleich zum menschlichen Urteil.

6. *Ausbildung und Supervision von Psychotherapeutinnen:* KI bietet vielversprechende Möglichkeiten für die Ausbildung und Supervision von Psychotherapeutinnen. Algorithmen könnten gezielte Rückmeldungen zu therapeutischen Fertigkeiten geben, etwa wie empathisch oder kongruent sich Therapeutinnen in Sitzungen verhalten. Durch die Analyse der Körpersprache, Mimik und Sprache könnte KI bspw. die Qualität der therapeutischen Beziehung objektiv messbar machen und darauf basierend Hinweise zur Optimierung geben. Diese objektiven Rückmeldungen könnten insbesondere in der Supervision genutzt werden, um blinde Flecken in der eigenen Wahrnehmung zu minimieren. Zudem könnten KI-gestützte Simulationen erstellt werden, die es angehenden Therapeutinnen ermöglichen, bestimmte Szenarien – wie den Umgang mit besonders schwierigen Klienten – gezielt zu üben. Erste Studien zeigen bereits einen positiven Einfluss auf die Qualität der klinischen Supervision (Cioffi et al., 2024).

7. *Spiegeltraining und Autismus-Spektrumstörungen:* Eine besonders interessante Methode für die Ausbildung ist das Spiegeltraining, bei dem Personen gebeten werden, die Emotionen einer Person in einem Video bewusst zu imitieren. Dieses Training, das sowohl die Wahrnehmung als auch das Ausdrucksvermögen schult, könnte durch KI erheblich optimiert werden. Die Algorithmen könnten detailliertes Feedback zu Genauigkeit und Intensität der Nachahmung geben und so die empathischen Fähigkeiten der Teilnehmenden gezielt stärken. Diese Methode ist nicht nur für angehende Psychotherapeuten relevant, sondern auch vielversprechend für die Arbeit mit Patientinnen auf dem Autismus-Spektrum, da sie hilft, emotionale Signale besser zu erkennen und zu verstehen – ein essenzieller Schritt, um deren zwischenmenschliche Kommunikation zu fördern. Gerade im Bereich Autismus wird zu dieser Methode dringend Forschung empfohlen (Banos et al., 2024).

8. *Nachweisbarkeit psychologischer Konzepte*: Ein faszinierender Anwendungsbereich wäre der Einsatz von KI zur Untersuchung klassischer psychoanalytischer Konzepte wie des *False Self* nach Winnicott (Ehrlich, 2021). Durch die Analyse parathymischer Phänomene – also die Diskrepanz zwischen dem, was jemand sagt, und dem, was er meint oder fühlt – könnten bestimmte psychologische Störungen genauer erforscht werden. Dies würde nicht nur die Diagnostik präzisieren, sondern auch ermöglichen, solche Aspekte gezielt in der Therapie anzusprechen.

9. *Anwendungen im forensischen und juristischen Bereich*: KI-gestützte Analyse könnten besonders auch im forensisch-psychiatrischen Bereich und anderen Kontexten wie Strafprozessen oder Glaubwürdigkeitsgutachten eine Rolle spielen. Ähnlich wie beim «Lügendetektor» wird man hier nicht erwarten können, dass die Maschine Entscheidungen trifft und «Krokodilstränen» oder rein «taktische Aussagen» identifiziert. Nichtsdestotrotz, wäre der Versuch der Identifikation solcher Fälle (v. a. auch retrospektiv durch Videoaufnahmen) ein lohnenswertes Unterfangen. Auch für Risikomodelle in der Suizidprävention zeigt KI bereits jetzt vielversprechende Ergebnisse (Lejeune et al., 2022). Der rechtliche Rahmen dieses Forschungsbereichs ist letztlich noch nicht vollständig geklärt, wenn auch hier ein gesteigertes öffentliches Interesse, besonders bei der Rückfallprophylaxe, gegeben scheint.

Grenzen KI-gestützter Analysen

Trotz ihres Potenzials stossen KI-gestützte Systeme in der Psychotherapieforschung und -praxis auf klare Grenzen. Ein zentraler Kritikpunkt ist die eingeschränkte Interpretierbarkeit vieler KI-Modelle. Insbesondere bei komplexen Algorithmen wie neuronalen Netzen bleibt oft unklar, wie genau ein Ergebnis zustande kommt. Diese Black-Box-Problematik erschwert nicht nur das Vertrauen in die Systeme, sondern auch deren Einsatz in klinischen Entscheidungsprozessen, wo Transparenz und Nachvollziehbarkeit unerlässlich sind (Carabantes, 2020).

Ein weiterer Aspekt ist die Abhängigkeit von der Interaktion zwischen Mensch und Maschine. Systeme, die auf einem *human in the loop*-(HITL)-Ansatz beruhen – also solche, bei denen Menschen die Ergebnisse der KI prüfen und interpretieren –, können die Grenzen der KI durch menschliche Expertise ausgleichen. Gleichzeitig erhöht dieser Ansatz den Arbeitsaufwand und verlangt hohe Kompetenzen seitens der Anwender. Im Gegensatz dazu setzen *machine in the loop*-(MITL)-Systeme verstärkt auf automatisierte Prozesse, bei denen Menschen lediglich überwachen oder nur im Ausnahmefall eingreifen (Demszky et al., 2023). Für den Bereich Psychotherapie ist es noch «ergebnisoffen», zu viele «klinische, ethische, rechtliche und praktische Fragen sind noch zu klären» (Kuhn et al., 2021).

Zudem bleibt die Generalisierbarkeit vieler Modelle ein Problem: KI-Systeme basieren auf spezifischen Datensätzen, die nicht unbedingt repräsentativ für die gesamte

Population sind, z. B. Altersbias (Kim et al., 2021). Dies birgt das Risiko, dass Ergebnisse für bestimmte Personengruppen nicht zutreffen oder sogar fehlerhaft sind, z. B. bei verschiedenen ethnischen Gruppen der seltenen psychischen Störungen (Mohammad, 2022). Hinzu kommen technische Herausforderungen wie das Risiko von *overfitting*, bei dem Modelle zwar perfekt auf Trainingsdaten abgestimmt sind, aber bei neuen Daten versagen. Schliesslich bleibt der rein menschliche Aspekt: Empathie, Intuition und die Fähigkeit, komplexe emotionale Nuancen wahrzunehmen, können von KI-Systemen bislang nicht vollständig erfasst oder ersetzt werden.

Datenschutzrechtliche und ethische Herausforderungen

Die potenziellen Vorteile neuer KI-Methoden in der Psychotherapie gehen unweigerlich mit erheblichen ethischen und datenschutzrechtlichen Herausforderungen einher. Erste Studien deuten bspw. darauf hin, dass KI bei der Prognose von selbstverletzendem Verhalten oder Suizidalität der menschlichen Einschätzung überlegen sein könnte (Li et al., 2024). Diese Entwicklungen könnten zu einer deutlichen Entlastung des klinischen Personals beitragen und die Qualität therapeutischer Entscheidungen erhöhen. Daraus ergibt sich die Forderung, Videoanalysen systematisch in der psychiatrischen Exploration einzusetzen. Gleichzeitig birgt diese Entwicklung erhebliche Risiken. Es besteht die Gefahr, dass digitale Prognosesysteme zu einem Selbstzweck werden und das menschliche Urteilsvermögen zunehmend der Maschine untergeordnet wird. In anderen Lebensbereichen, in denen Algorithmen bereits dominieren, hat sich gezeigt, wie schnell ein solches Ungleichgewicht entstehen kann. Kritiker warnen in diesem Zusammenhang vor einem schleichenden Übergang zu einem «digitalen Totalitarismus» (Zuboff, 2019), in dem automatisierte Systeme nicht nur Entscheidungen dominieren, sondern auch persönliche Freiheiten bedrohen.

Dieses Spannungsfeld wird in der Psychotherapie durch ein weiteres ethisches Dilemma verstärkt: Einerseits erfordert die Verbesserung von KI-Modellen umfassende und detaillierte Datensätze, was die systematische Videoaufzeichnung von Therapiesitzungen nahelegt. Solche Daten könnten die Präzision und Effektivität von KI-Systemen weiter steigern. Andererseits ist der Therapieraum ein besonders schützenswerter Ort, der Vertraulichkeit und die ärztliche Schweigepflicht gewährleisten muss – beides zentrale Säulen einer wirksamen und vertrauensvollen Therapie. Die Herausforderung besteht darin, eine Balance zu finden, die den technologischen Fortschritt ermöglicht, ohne die ethischen Grundwerte und die Autonomie der Patientinnen zu gefährden.

Gerade in einem so sensiblen Feld wie der Psychotherapie sind die ethischen und datenschutzrechtlichen Herausforderungen daher enorm. Der Einsatz von KI erfordert höchste Standards in der Verarbeitung und Speicherung personenbezogener Daten, die oft emotionaler oder physiologischer Natur sind. Sichere Datenräume

(*secure data rooms*) oder *trusted research environments* (TRE) bieten hier eine wichtige Lösung (Okoniewski et al., 2024). Eine andere Möglichkeit sind so genannte *algorithm-to-data*-Lösungen, bei denen nicht die sensiblen Daten selbst (z. B. Videos) geteilt werden, sondern der Algorithmus die Daten aufsucht und lediglich aggregierte Analyseergebnisse zurückgibt. Diese Ansätze ermöglichen es, dass Algorithmen auf Daten zugreifen, ohne die Privatsphäre der Patienten zu gefährden. Sie schaffen geschützte Umgebungen, die garantieren, dass sensible Informationen nur von autorisierten Personen und Systemen verarbeitet werden. Solche technischen Lösungen könnten ein wichtiger Schritt sein, um die Balance zwischen ethischen Anforderungen und technologischem Fortschritt zu wahren.

Zusätzlich stellt die Einzigartigkeit bestimmter Daten, wie etwa mimischer Verhaltensweisen, eine besondere Herausforderung in diesem Bereich dar. Diese basieren auf biometrischen Daten und könnten ähnlich wie ein Fingerabdruck zur Identifikation von Personen genutzt werden, was erhebliche Datenschutzrisiken birgt (Czarnecki, 2021). Um solche Risiken zu minimieren, bedarf es weiterführender Forschung darüber, welche Daten für Forschungszwecke sicher geteilt werden können und wie Datenmanagementpläne sowie Open-Data-Strategien gestaltet werden sollten. Nur durch eine enge Abstimmung zwischen Forschung und Datenschutzrecht lassen sich praktikable Lösungen entwickeln, die wissenschaftlichen Fortschritt ermöglichen und gleichzeitig die Privatsphäre der Patientinnen schützen.

Abschliessend ist es wichtig, den Einsatz von KI in der Psychotherapie immer kritisch zu hinterfragen und sicherzustellen, dass sie nicht als Ersatz, sondern als Ergänzung zum menschlichen Urteil dient. Nur durch den verantwortungsvollen Einsatz von Technologien wie *secure data rooms*, einer sinnvollen Regulierung und kontinuierlicher Reflexion kann sichergestellt werden, dass der Nutzen von KI-Systemen das Wohl der Patienten fördert, ohne ihre Autonomie oder Privatsphäre zu gefährden, und gleichzeitig Innovation nicht blockiert.

Fazit und Ausblick

Die Integration von KI in die Psychotherapie markiert einen Wendepunkt in der Forschung und Praxis. Die Möglichkeit, hochauflösende Daten über Körpersprache, Mimik, Sprache und emotionale Zustände systematisch zu erfassen und auszuwerten, eröffnet völlig neue Perspektiven. Diese *high resolution psychotherapy* bietet die Chance, psychotherapeutische Prozesse in bisher unbekannter Präzision zu analysieren, langfristig zu dokumentieren und sowohl die Forschung als auch die Supervision von Therapeutinnen grundlegend zu verändern.

Gleichzeitig zeigt sich, dass dieser Fortschritt ohne klare ethische und methodische Leitplanken nur begrenzt realisierbar ist. Die Etablierung neuer Standards wird essenziell sein, um die Integrität der Therapie zu wahren und den Patientinnen in den Mittelpunkt zu stellen. Dies

erfordert nicht nur technologische Innovation, sondern auch gesellschaftlichen Konsens über den verantwortungsvollen Einsatz solcher Systeme.

In der Zukunft könnten KI-gestützte Methoden nicht nur helfen, psychologisch schwergreifbares wie Authentizität oder emotionale Resonanz messbar zu machen, sondern auch die Interaktion zwischen Patientin und Therapeut auf neue Weise bereichern. Dabei bleibt die zentrale Frage: Wie können wir die datengetriebene Zukunft der Psychotherapie gestalten, ohne die fundamentalen Werte wie Empathie, Vertrauen und Menschlichkeit aus den Augen zu verlieren? Der Schlüssel liegt in der harmonischen Verbindung von technologischem Fortschritt mit den bewährten Prinzipien therapeutischer Praxis.

Danksagung

Wir danken Frau (cand. psych.) Nora Balint herzlich für ihre wertvolle Unterstützung bei den Filmaufnahmen zur Synchronizität von Emotionen.

Literatur

- Agrawal, A. & Mittal, N. (2020). Using CNN for facial expression recognition: A study of the effects of kernel size and number of filters on accuracy. *The Visual Computer*, 36(2), 405–412.
- Altman, N. & Krzywinski, M. (2018). The curse(s) of dimensionality. *Nat Methods*, 15(6), 399–400.
- Arabskyy, Y. et al. (2021). *Dialectal Speech Recognition and Translation of Swiss German Speech to Standard German Text: Microsoft's Submission to SwissText 2021* arXiv. <https://doi.org/10.48550/arXiv.2106.08126>
- Arik, S.Ö. et al. (2017). Deep voice: Real-time neural text-to-speech. *International conference on machine learning*, 195–204.
- Bajwa, M.N. et al. (2020). Confident classification using a hybrid between deterministic and probabilistic convolutional neural networks. *IEEE Access*, 8, 115476–115485.
- Bänninger-Huber, E. & Steiner, F. (1992). Identifying Microsequences: A New Methodological Approach to the Analysis of Affective Regulatory Processes. In M. Leuzinger-Bohleber et al. (Hg.), *«Two Butterflies on My Head ...»* (S. 257–276). Springer.
- Bänninger-Huber, E. (2021). Prozesse der Emotionsregulierung in psychoanalytischen Langzeitpsychotherapien. *Psychotherapie Forum*, 25(1–2), 80–87. <https://doi.org/10.1007/s00729-021-00172-7>
- Bänninger-Huber, E. & Widmer, C. (2020). Affective relationship patterns and psychotherapeutic change. What the face reveals. *Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*, 456–468.
- Banos, O. et al. (2024). Sensing technologies and machine learning methods for emotion recognition in autism: Systematic review. *International Journal of Medical Informatics*, 105469.
- Beebe, B. (1982). Micro-Timing in Mother-Infant Communication. In M.R. Key (Hg.), *Nonverbal Communication Today* (S. 169–196). de Gruyter.
- Benecke, C. & Krause, R. (2005). Initiales mimisch-affektives Verhalten und Behandlungszufriedenheit in der Psychotherapie von Patientinnen mit Panikstörungen/Initial Affective Facial Behaviour and Outcome Satisfaction in the Psychotherapy of Patients with Panic Disorder. *Zeitschrift für Psychosomatische Medizin und Psychotherapie*, 51(4), 346–359. <https://doi.org/10.13109/zptm.2005.51.4.346>
- Berisha, V. et al. (2021). Digital medicine and the curse of dimensionality. *NPJ digital medicine*, 4(1), 153.

- Buchheim, A. & Benecke, C. (2007). Mimisch-affektives Verhalten bei Patientinnen mit Angststörungen während des Adult-Attachment-Interviews: Eine Pilotstudie. *PPmP*, 57(8), 343–347. <https://doi.org/10.1055/s-2006-952030>
- Cabitza, F. et al. (2022). The unbearable (technical) unreliability of automated facial emotion recognition. *Big Data & Society*, 9(2), 20539517221129549. <https://doi.org/10.1177/20539517221129549>
- Carabantes, M. (2020). Black-box artificial intelligence: An epistemological and critical analysis. *AI & society*, 35(2), 309–317.
- Chattopadhyay, A. & Lu, T.-P. (2019). Gene-gene interaction: The curse of dimensionality. *Annals of translational medicine*, 7(24). <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6989881>
- Chen, Z. et al. (2023). *Empowering Psychotherapy with Large Language Models: Cognitive Distortion Detection through Diagnosis of Thought Prompting*. arXiv. <https://doi.org/10.48550/arXiv.2310.07146>
- Cioffi, V. et al. (2024). *Can AI Technologies Support Automating Clinical Supervision? Assessing the Potential of ChatGPT*.
- Czarnocki, J. (2021). Will new definitions of emotion recognition and biometric data hamper the objectives of the proposed AI Act? *2021 International Conference of the Biometrics Special Interest Group (BIOSIG)*, 1–4.
- Dael, N. et al. (2012). The Body Action and Posture Coding System (BAP): Development and Reliability. *Journal of Nonverbal Behavior*, 36(2), 97–121. <https://doi.org/10.1007/s10919-012-0130-0>
- Darwin, C. (1872). *The expression of the emotions in man and animals*. Oxford UP.
- Darboven, J. et al. (2024). Hürdenlauf zur Psychotherapie: Was hält Menschen davon ab, eine Psychotherapie in Anspruch zu nehmen? *24. Nachwuchswissenschaftler:innenkonferenz*, 3, 117–122. <https://monami.hs-mittweida.de/files/15378/Darboven.pdf>
- De Waal, F. (2014). *Our inner ape: The best and worst of human nature*. Granta Books.
- Demszky, D. et al. (2023). Using large language models in psychology. *Nature Reviews Psychology*, 2(11), 688–701.
- Duncan, L.E. et al. (2019). How genome-wide association studies (GWAS) made traditional candidate gene studies obsolete. *Neuropsychopharmacology*, 44(9), 1518–1523.
- Ehrlich, R. (2021). Winnicott's Idea of The False Self: Theory as Autobiography. *Journal of the American Psychoanalytic Association*, 69(1), 75–108. <https://doi.org/10.1177/00030651211001461>
- Ekman, P. & Friesen, W.V. (1971). Constants across cultures in the face and emotion. *Journal of personality and social psychology*, 17(2), 124.
- Ekman, P. & Friesen, W.V. (1974). Nonverbal behavior and psychopathology. *The psychology of depression: Contemporary theory and research*, 3–31.
- Ekman, P. & Friesen, W.V. (1978). Facial action coding system. *Environmental Psychology & Nonverbal Behavior*. <https://psycnet.apa.org/doiLanding?doi=10.1037/t27734-000>
- Eskimez, S.E. et al. (2016). Emotion classification: How does an automated system compare to naive human coders? *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2274–2278.
- Fiske, A. et al. (2019). Your robot therapist will see you now: Ethical implications of embodied artificial intelligence in psychiatry, psychology, and psychotherapy. *Journal of medical Internet research*, 21(5), e13216.
- Fürer, L. et al. (2020). Supervised speaker diarization using random forests: A tool for psychotherapy process research. *Frontiers in psychology*, 11, 1726.
- Han, B. et al. (2020). Toward unbiased facial expression recognition in the wild via cross-dataset adaptation. *IEEE Access*, 8, 159172–159181.
- Jaiswal, S. & Nandi, G.C. (2020). Robust real-time emotion detection system using CNN architecture. *Neural Computing and Applications*, 32(15), 11253–11262. <https://doi.org/10.1007/s00521-019-04564-4>
- Kim, E. et al. (2021). Age Bias in Emotion Detection: An Analysis of Facial Emotion Recognition Performance on Young, Middle-Aged, and Older Adults. *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, 638–644. <https://doi.org/10.1145/3461702.3462609>
- Kuhn, E. et al. (2021). Psychotherapie mit einer autonomen Künstlichen Intelligenz. Ethische Chancen und Herausforderungen 1. *Psychiatrische Praxis*, 48(S 01), S26–S30. <https://doi.org/10.1055/a-1369-2938>
- Lejeune, A. et al. (2022). Artificial intelligence and suicide prevention: A systematic review. *European psychiatry*, 65(1), e19.
- Li, X. et al. (2024). Exploring the Potential of Artificial Intelligence in Adolescent Suicide Prevention: Current Applications, Challenges, and Future Directions. *Psychiatry*, 87(1), 7–20. <https://doi.org/10.1080/00332747.2023.2291945>
- Lin, B. et al. (2022). *SupervisorBot: NLP-Annotated Real-Time Recommendations of Psychotherapy Treatment Strategies with Deep Reinforcement Learning*. arXiv. <https://doi.org/10.48550/arXiv.2208.13077>
- McNeish, D.M. (2014). Modeling sparsely clustered data: Design-based, model-based, and single-level methods. *Psychological methods*, 19(4), 552.
- Mehrabian, A. & Ferris, S.R. (1967). Inference of attitudes from non-verbal communication in two channels. *Journal of consulting psychology*, 31(3), 248.
- Merten, J. et al. (1996). Emotional Experience and Facial Behavior During the Psychotherapeutic Process and its Relation to Treatment Outcome: A Pilot Study. *Psychotherapy Research*, 6(3), 198–212. <https://doi.org/10.1080/10503309612331331708>
- Miner, A.S. et al. (2020). Assessing the accuracy of automatic speech recognition for psychotherapy. *NPJ digital medicine*, 3(1), 82.
- Mohammad, S.M. (2022). Ethics sheet for automatic emotion recognition and sentiment analysis. *Computational Linguistics*, 48(2), 239–278.
- Najjar, R. (2023). Redefining radiology: A review of artificial intelligence integration in medical imaging. *Diagnostics*, 13(17), 2760.
- Orrù, G. et al. (2020). Machine learning in psychometrics and psychological research. *Frontiers in psychology*, 10, 2970.
- Okoniewski, M.J. et al. (2024). Leonhard Med, a trusted research environment for processing sensitive research data. *Journal of Integrative Bioinformatics*, 21(3), 20240021. <https://doi.org/10.1515/jib-2024-0021>
- Park, S. et al. (2021). The analysis of emotion authenticity based on facial micromovements. *Sensors*, 21(13), 4616.
- Ramseyer, F.T. (2020). Motion energy analysis (MEA): A primer on the assessment of motion from video. *Journal of counseling psychology*, 67(4), 536.
- Ranjbarzadeh, R. et al. (2023). Brain tumor segmentation of MRI images: A comprehensive review on the application of artificial intelligence tools. *Computers in biology and medicine*, 152, 106405.
- Reis, D. & Friese, M. (2022). The Myriad Forms of p-Hacking. In W. O'Donohue et al. (Hg.), *Avoiding Questionable Research Practices in Applied Psychology* (S. 101–121). Springer.
- Riffer, F. et al. (2021). Geschlechtsspezifische Unterschiede in der psychotherapeutischen Versorgung: Aktuelle Daten aus Österreich. *Psychotherapeut*, 66(6), 511–517. <https://doi.org/10.1007/s00278-021-00523-4>
- Rinn, W.E. (1984). The neuropsychology of facial expression: A review of the neurological and psychological mechanisms for producing facial expressions. *Psychological bulletin*, 95(1), 52.
- Saxen, F. et al. (2017). Real vs. fake emotion challenge: Learning to rank authenticity from facial activity descriptors. *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 3073–3078. https://openaccess.thecvf.com/content_ICCV_2017_workshops/w44/html/Saxen_Real_vs._Fake_ICCV_2017_paper.html
- Singh, A.K. et al. (2022). Real-Time Human Pose Detection and Recognition Using MediaPipe. In V.S. Reddy et al. (Hg.), *Soft Computing and Signal Processing* (S. 145–154). Springer Nature.

- Steppan, M. et al. (2024). Machine Learning Facial Emotion Classifiers in Psychotherapy Research: A Proof-of-Concept Study. *Psychopathology*, 57(3), 159–168.
- Tal, S. et al. (2024). Reactivity and stability in facial expressions as an indicator of therapeutic alliance strength. *Psychotherapy Research*, 1–15. <https://doi.org/10.1080/10503307.2024.2311777>
- Tronick, E. et al. (1978). The infant's response to entrapment between contradictory messages in face-to-face interaction. *Journal of the American Academy of Child psychiatry*, 17(1), 1–13.
- Tyrrrell, J. et al. (2021). Genetic predictors of participation in optional components of UK Biobank. *Nature communications*, 12(1), 886.
- Zimmermann, R. et al. (2021). Movement synchrony in the psychotherapy of adolescents with borderline personality pathology. A dyadic trait marker for resilience? *Frontiers in psychology*, 12, 660516.
- Zuboff, S. (2019). Surveillance Capitalism and the Challenge of Collective Action. *New Labor Forum*, 28(1), 10–29. <https://doi.org/10.1177/1095796018819461>

High resolution psychotherapy research through AI-supported process analysis A paradigm shift in psychotherapy research

Abstract: Psychotherapy research is undergoing a fundamental paradigm shift driven by the use of artificial intelligence (AI). While Paul Ekman's work on emotion recognition was based on the observation and manual coding of facial expressions, AI now enables automated, multimodal analysis of speech, facial expressions, gestures, and emotional dynamics. These technologies allow for the near-complete recording, documentation, and analysis of entire therapy sessions in real time for the first time. This paper examines the transition from Ekman's methodological approach to AI-supported «total data collection» and analyzes the scientific and clinical implications. Special attention is given to the opportunities arising from high-resolution real-time analysis of therapeutic processes, as well as the risks concerning data privacy and ethical standards. Finally, the discussion addresses how anonymization and strict security measures can ensure the responsible use of sensitive therapy data, unlocking the full potential of these technologies.

Keywords: artificial intelligence, epistemology, emotion recognition, affective computing, clinical supervision, psychotherapy, psychiatry, new methods

Ricerca sulla psicoterapia ad alta risoluzione attraverso l'analisi dei processi supportata dall'IA Un cambiamento di paradigma nella ricerca in psicoterapia

Riassunto: La ricerca in psicoterapia sta vivendo un fondamentale cambiamento di paradigma grazie all'uso dell'intelligenza artificiale (IA). Mentre il lavoro di Paul Ekman sul riconoscimento delle emozioni si basava sull'osservazione e sulla codifica manuale delle espressioni facciali, l'IA consente oggi l'analisi automatizzata e multimodale del parlato, delle espressioni

facciali, dei gesti e delle dinamiche emotive. Per la prima volta, queste tecnologie consentono la registrazione, la documentazione e l'analisi quasi completa di intere sessioni terapeutiche in tempo reale. L'articolo esamina la transizione dall'approccio metodologico di Ekman alla «elicitazione totale» supportata dall'IA e analizza le implicazioni scientifiche e cliniche. Viene prestata particolare attenzione alle opportunità derivanti dall'analisi in tempo reale ad alta risoluzione dei processi terapeutici, nonché ai rischi relativi alla protezione dei dati e agli standard etici. Infine, viene discusso come l'anonimizzazione e le rigorose misure di sicurezza possano garantire una gestione responsabile dei dati terapeutici sensibili, al fine di sfruttare appieno il potenziale di queste tecnologie.

Parole chiave: intelligenza artificiale, epistemologia, riconoscimento delle emozioni, elaborazione dei dati affettivi, monitoraggio clinico, psicoterapia, psichiatria, nuovi metodi

Biografische Notiz

Martin Steppan hat Psychologie (mit Pharmazie im Nebenfach) studiert. Er promovierte an der Universität Innsbruck zu Suchterkrankungen sowie an der Universität St. Andrews zur Verhaltensgenetik, insb. zum Einfluss der Pubertät auf psychische Störungen. Er war langjährig am Münchner IFT Institut für Therapieforschung sowie an den Universitären Psychiatrischen Kliniken Basel tätig, mit einem Schwerpunkt auf Persönlichkeitsstörungen. Seit 2021 ist er Senior Research and Teaching Assistant an der Fakultät für Psychologie der Universität Basel, wo er Psycholog:innen in Forschungsmethodik unterrichtet. Sein besonderes Interesse gilt methodischen Innovationen, insb. in der psychologischen Diagnostik und Psychotherapieforschung. Dabei arbeitet er an neuen Ansätzen zur KI-gestützten Emotionsanalyse sowie an innovativen psychometrischen Verfahren.

Marc Birkhölzer hat in Gießen Medizin studiert und lebt seit 2013 in der Schweiz. Seit 2010 forscht er zum Thema Jugendliche mit Persönlichkeitsstörungen, hier insb. zum Thema Früherkennung und Frühintervention. Er hat seine Psychotherapieweiterbildung in systemisch-analytischer Kinder-, Jugend- und Familientherapie absolviert und ist Facharzt für Kinder- und Jugendpsychiatrie und -psychotherapie. Er ist seit Oktober 2021 Oberarzt der Jugendforensischen Ambulanz (JAM) der Klinik für Forensik, UPK Basel, Forschungsgruppenleiter der EARLY-Studie und seit August 2023 Mitglied der Institutsleitung des Instituts für Kinder-, Jugend- und Familientherapie (KIF) Luzern. Sein besonderes Interesse gilt der Persönlichkeitsentwicklung, der Adoleszenz, dissozialem Verhalten und der Erforschung des Potenzials psychoaktiver Substanzen in der Psychotherapie.

Kontakt

martin.steppan@unibas.ch