Vorwissenschaftliche Arbeit

über das Thema:

Korrelation zwischen Gehirnforschung und Künstlicher Intelligenz

verfasst von:
Larissa Arthofer
Klasse: 8B

Betreuerin: Mag. Sonja Koger

Abgabedatum: 15/02/2024

Bundesrealgymnasium 19

Krottenbachstraße 11-13, 1190 Wien

Abstract

This paper delves into the captivating correlations between two groundbreaking domains of science: brain research and artificial intelligence (AI). The primary objective is to uncover the structural and functional congruencies between AI models and the human brain, as well as to deduct interdisciplinary ethical and legal ramifications. Employing an empirical approach, this research integrates a hands-on exploration through a custom AI model, expert interviews, a tailored legal draft, and extensive literature review. Three core research questions underpin this study.

First, it explores how the human brain's operating principles are reflected in AI algorithms, highlighting the enhancement of AI efficiency and effectiveness through neuroscientific insights.

Second, the contribution of brain research to AI's design and optimization for pattern recognition is examined, particularly through a convolutional neural network case study that reveals parallels with human neural processes.

Third, this paper addresses the ethical and legal implications of the convergence of AI and brain research, focusing on bias mitigation. A legal framework is crafted to guide the regulation of bias in AI, emphasizing the integration of legal considerations informed by neurological research.

Vorwort

Entstehungsgeschichte

Künstliche Intelligenz ist das Gebiet, welches meine Neugier und Leidenschaft in diesem Jahr entfacht hat. Erst vor wenigen Monaten begann mein Abenteuer in diese faszinierenden Welt, und doch fühlt es sich an, als hätte ich bereits eine Reise mit unzähligen Erkenntnissen und Lernerfolgen hinter mir.

Die grenzenlosen Möglichkeiten, die Künstliche Intelligenz für die Zukunft in Aussicht stellt, haben mich in ihren Bann gezogen. Dabei war es nicht nur die Technologie selbst, die mich faszinierte, sondern vor allem die erstaunliche Ähnlichkeit zur Funktionsweise unseres Gehirns. Die Art und Weise, wie KI lernt, die Strukturen, die neuronalen Netzwerke – all das sind künstliche Nachbildungen unseres biologischen Denkorgans. Der entscheidende Unterschied? Eine unschlagbare Rechenleistung, die die Türen zu unendlichen Möglichkeiten in der Zukunft öffnet.

Diesen Sommer, im Jahr 2023, hatte ich das Privileg, noch tiefer in diese aufregende Welt einzutauchen. Ich begann bei Women in AI Austria zu arbeiten, tauchte im Rahmens meines Praktikums in die Welt des IT/IP - und KI-Rechts bei der DORDA Rechtsanwaltskanzlei ein und sammelte wertvolle Erfahrungen auf zahlreichen Festivals und Veranstaltungen, bei denen ich von den klügsten Köpfen auf diesem Gebiet lernen konnte. Zusätzlich bekam ich die Gelegenheit, den Harvard Online Course "Introduction to Python" zu absolvieren, der mir die Grundlagen des Programmierens und die Implementierung von KI näherbrachte.

Dieses Wissensgebiet eröffnete mir völlig neue Horizonte und Fähigkeiten. Kurzzeitig schwankte sogar meine lebenslange Entscheidung Recht studieren und ich fragte mich, ob ich im Herbst mein Studium der Rechtswissenschaften beginnen oder dem Beispiel meiner neuen, inspirierenden Mentoren folgen und mich tiefer in die Welt der Künstlichen Intelligenz stürzen sollte.

Die schier endlosen Möglichkeiten, die KI mit sich bringt, faszinieren mich nach wie vor zutiefst, und mein Ziel in dieser vorwissenschaftlichen Arbeit ist es, herauszufinden, wie eng die Verbindungen zwischen KI und unserem Gehirn tatsächlich sind. Durch meine ersten praktischen Erkundungen dieser neuen Welt und mit dem frischen Wissen, das ich gesammelt habe, möchte ich meine Erkenntnisse und meine Begeisterung mit der Welt teilen.

Diese Arbeit widmet sich der Erkundung der Parallelen zwischen KI und unserem Denkorgan und stellt sich dabei die Frage, inwiefern sich unser Gehirn in der Welt der Künstlichen Intelligenz widerspiegelt.

Danksagung

Ich möchte mich bei all den unglaublichen Menschen bedanken, die mich auf dieser Reise begleitet und inspiriert haben. Ihre Unterstützung und unerschütterlicher Glaube an mich haben diese Arbeit erst möglich gemacht.

Ein riesiges Dankeschön gilt meiner Betreuerin, Mag. Sonja Koger. Ihre Unterstützung und Geduld waren von unschätzbarem Wert.

Ein besonderer Dank geht an das fantastische Team von Women in AI Austria. Mag. Carina Zehetmaier, E.MA die mich nicht nur inspiriert und als Mentorin begleitet, sondern mich auch als das "Prodigy" (DE: Wunderkind) bei Women in AI Austria aufgenommen und von all ihren Keynotes und Erfahrungen profitieren hat lassen. Vielen Dank auch an Mag. Alexandra Ciarnau, die mich voller Weisheit und Inspiration auf meinem Weg ermutigt hat, und mir die Tür zu meinem Praktikum bei der DORDA Rechtsanwaltskanzlei geöffnet hat, wo ich unglaublich viel von ihr lernen konnte, und mir ein Einblick in die Schnittstelle von KI und IT-Recht ermöglicht wurde.

Ein Dank geht auch an Mag. Alexander Zehetmaier, Mag. Natascha Windholz und Dr. DI. Mag. Isabella Hinterleitner, dessen Interviews und Expertise in den Bereichen Programmieren, Datenschutz und KI, eine riesige Bereicherung für meine Arbeit waren. Ihr habt mich ermutigt und inspiriert, immer weiterzumachen. Noch dazu bedanke ich mich bei Mag. Juliane Lörincz, die meine Leidenschaft für wissenschaftliches Arbeiten und die Neurowissenschaften entfacht hat.

Meiner Familie, insbesondere Sonja, Maria, Lutz, und Benedikt, danke ich von Herzen. Sie waren immer an meiner Seite, haben mich ermutigt und wurden nie müde von meiner endlosen KI-Begeisterung und all meinen neuen Erkenntnissen und Entdeckungen zu hören. Ihr seid meine größte Inspiration.

Zudem wäre diese Arbeit nicht ohne der Open Source Community möglich gewesen. Zuletzt möchte ich mich bei den Organisatoren der Fachpreise bedanken, bei denen ich meine Arbeit eingereicht habe. Die Aussicht auf Anerkennung haben mir eine riesige Ressourcenmenge an Energie, Kraft und Begeisterung geschenkt.

Ihre Unterstützung hat mir den Weg geebnet, und dafür bin ich unendlich dankbar.

1. September 2023 Larissa Arthofer 1020 Wien

Inhaltsverzeichnis

Gender-Hinweis

Die in dieser VWA verwendeten Personenbezeichnungen beziehen sich immer gleichermaßen auf weibliche und männliche Personen. Auf eine Doppelnennung und gegenderte Bezeichnungen wird zugunsten einer besseren Lesbarkeit verzichtet.[1]

[1] Um die Verständlichkeit zusätzlich zu erhöhen, befindet sich auf Seite 36-39 ein Glossar, in dem alle wichtigen Fachbegriffe nachgeschlagen werden können.

1 Einleitung

17 Jahre, so viel Zeit habe ich bisher damit verbracht, meine Hirnkapazitäten auszubauen, um zu verstehen, zu lesen, zu schreiben, zu sprechen und zu rechnen.
Sekundenbruchteile – das ist die Dauer, die eine Künstliche Intelligenz benötigt, um 200 Billiarden Berechnungen durchzuführen, wofür ein Mensch Milliarde Jahre brauchen würde.[2]
4 Stunden, und sie wird vom Schachanfänger zum Schachweltmeister.[3]
67 Stunden, und sie ist imstande, komplexe Proteinstrukturen vorherzusagen, ein Rätsel an dem menschliche Forschungsteams Jahrzehnte lang gearbeitet haben.[4]

100 Milliarden Neuronen ersetzt durch weniger als 1 Million künstlichen Neuronen.

Wenn man sich vorstellt, seine gesamte Gehirnleistung exponentiell um die Potenz drei zu erweitern, dazu noch das geballte Wissen des gesamten Internets aufzunehmen, und das alles, während man einen Energydrink trinkt. Das ist die Rechenleistung der KI – eine Leistung, die unser menschliches Gehirn in den Schatten stellt.

Es stellt sich die Frage, worin der wirkliche Unterschied liegt. Ist Künstliche Intelligenz lediglich ein ausgeklügelter, rechenstarker Nachbau unseres Gehirns oder verbirgt sich doch mehr dahinter?

Die folgende Arbeit widmet sich der Erforschung der grundlegenden Verarbeitungsweisen und Synergien zwischen Künstlicher Intelligenz und unserem Gehirn. Die Erkenntnisse basieren nicht nur auf einer umfangreichen Literaturanalyse und Expertinnen Interviews, sondern auch auf der Analyse des Fallbeispiels einer selbstprogrammierten KI. Dabei geht es darum, die Korrelation zwischen der Funktionsweise unserer neuronalen Netze und deren Einfluss auf die Struktur und Weiterentwicklung der KI, mit einem besonderen Fokus auf die Mustererkennung, zu erforschen. Doch das ist noch nicht alles.

Neben den technischen und neurologischen Aspekten beleuchtet diese Arbeit auch interdisziplinäre Ansätze wie die ethischen und rechtlichen Perspektiven, die sich aus dieser neurologischen Konvergenz ergeben. Diese kritische Auseinandersetzung hebt die Relevanz und Aktualität des Themas hervor, da sie auf aktuellen gesellschaftlichen Problemen basiert und realistische Lösungsansätze in der Form eines Gesetzesentwurfes skizziert. Die Erforschung des Gehirns und die Entwicklung von Künstlicher Intelligenz sind zwei der bahnbrechendsten Bereiche der Wissenschaft unserer Zeit. Die Arbeit bietet eine Ausschau in die Zukunft, und wie diese außergewöhnliche Konvergenz die Welt, wie wir sie kennen, verändert und prägt.

---

[2] vgl. SAP, https://www.sap.com/austria/products/artificial-intelligence/what-is-artificial-intelligence.html#:~:text=Starke%20künstliche%20Intelligenz%20(starke%20KI)&text=Der%20Summit%20Supercomputer%20ist%20einer,dafür%20eine%20Milliarde%20Jahre%20brauchen., 9.11.2023

[3] vgl. Breithut J, https://www.spiegel.de/netzwelt/web/google-ki-alphazero-meistert-schach-und-go-a-1182395.html, 02.11.2023

[4] vgl. Cheng S., u.a, https://arxiv.org/pdf/2203.00854.pdf#:~:text=We%20successfully%20scaled%20the%20AlphaFold,to%20significtant%20cost%20sav%2D%20ings., 02.11.2023

Diese VWA bietet insgesamt einen faszinierenden Einblick in die Welt der KI und der Neurowissenschaften, die uns in den kommenden Jahren und Jahrzehnten zweifellos weiterhin fesseln und beeinflussen wird. Sie zeugt von meiner Leidenschaft für dieses Thema und meiner Entschlossenheit, die Innovationen und Herausforderungen dieses aufstrebenden Forschungsfelds zu erkunden.

2 Thematische Einführung

Um dieses spannende Thema aus verschiedenen Blickwinkeln zu betrachten habe ich die folgenden drei Leitfrage gewählt, die in den folgenden Kapitel untersucht werden.

1. *Auf welche Weise spiegeln sich die Arbeitsweisen des menschlichen Gehirns in den Funktionsprinzipien von KI-Algorithmen wider?*

2. *Wie tragen Erkenntnisse aus der Hirnforschung zur Struktur und Optimierung von KI-Technologien zur Mustererkennung bei?*

3. *Welche ethischen und rechtlichen Überlegungen ergeben sich aus der wachsenden Konvergenz von KI und neurologischem Verständnis?*

2.1 Grundlagen

Eine Definition von KI lautet: *„Künstliche Intelligenz ist die Fähigkeit einer Maschine, menschliche Fähigkeiten wie logisches Denken, Lernen, Planen und Kreativität zu imitieren.“*[5]
Diese Definition ist vielschichtig und umfasst selbst in den neuesten juristischen Texten wir dem EU AI Act, eine Vielzahl von technologischen Ansätzen basierend auf maschinellen Lernen, Deep Learning, logik- und wissensbasierten Systemen sowie statistischen Methoden.[6]
In dieser Arbeit liegt ein besonderer Fokus auf höhere kognitive Funktionen wie Denken, Lernen, Sprache und Bewusstsein. Einen Einstieg zu dem Thema bildet eine kurze Erklärung zu den Arbeitsweisen und den Aufbau von KI und den neurologischen Mechanismen des menschlichen Gehirns.

Künstliche Intelligenz ist ein Teilgebiet der Informatik, und wird üblicherweise mit den Programmiersprachen Python oder R programmiert. In der Regel funktioniert KI so, dass sie Input-Daten erhält, welche sie mit einem vielschichtigen, versteckten Algorithmus (hidden layers) verarbeitet und anschließend Outputs ergibt. Die folgende Abbildung die mit adaptiertem Code von Stack Exchange in Latex erstellt wurde, stellt eine vereinfachte Darstellung der Arbeitsweise einer künstlichen Intelligenz dar:

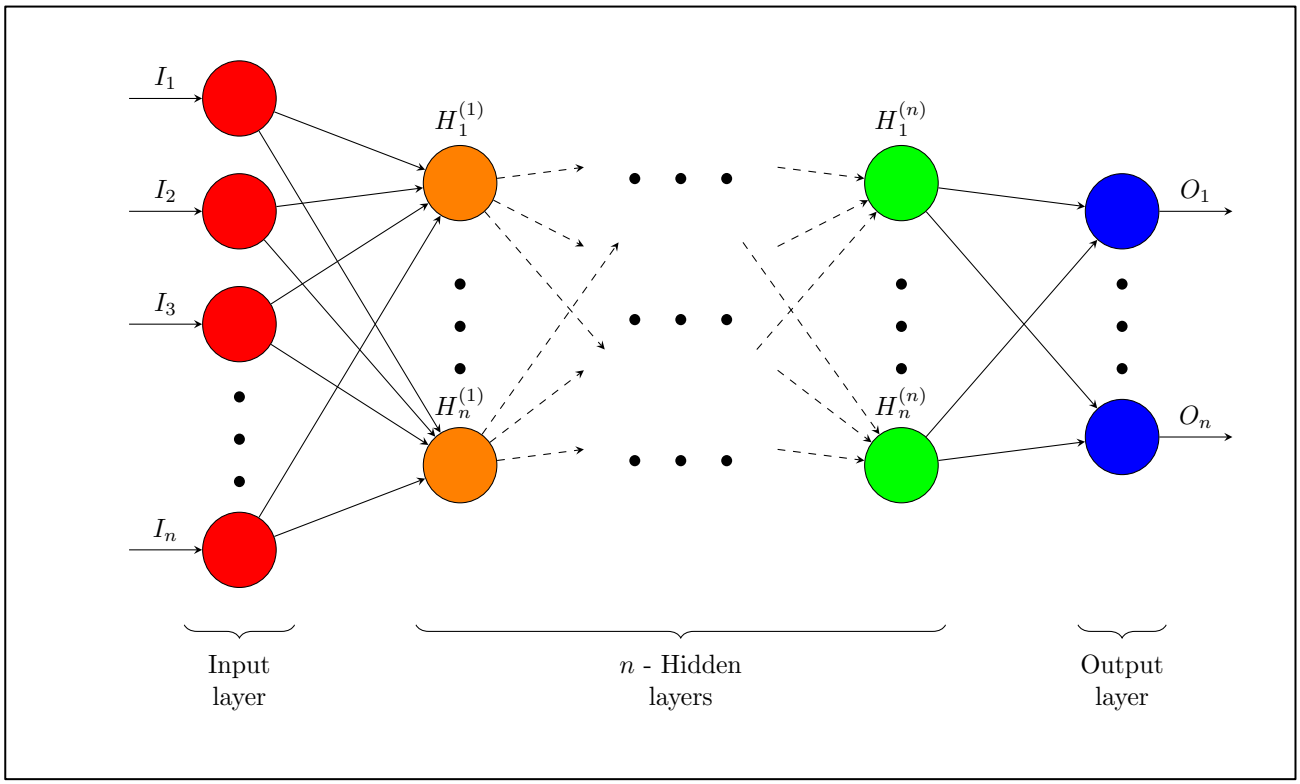


Abbildung 1: Architektur eines künstlichen neuronalen Netzwerks [7]

[5] Europäisches Parlament, https://www.europarl.europa.eu/news/de/headlines/society/20200827STO85804/was-ist-kunstliche-intelligenz-und-wie-wird-sie-genutzt (10. August 2023)
[6] vgl. Wide Angle Analytics, https://wideangle.co/blog/ai-regulation-eu-ai-act (13. August 2023)
[7] Eigene Darstellung (siehe Anhang 1)

Künstliche Intelligenz (KI) arbeitet mit künstlichen neuronalen Netzwerken. Diese senden elektrische Signale über spezielle Graphikprozessoren (GPUs), welche mehrere Neuronen simultan berechnen können (Parallelverarbeitung), was zu einer optimierten Leistung führt.[8] KI unterscheidet sich von anderen technologischen Systemen durch ihre Autonomität und Adaptivität. Darunter versteht man, dass sie sich selbstständig optimiert, aus Erfahrungen lernt und autonom Vorhersagen und Entscheidungen treffen kann.

Eine Besonderheit von künstlichen neuronalen Netzwerken stellt dar, dass sie üblicherweise nicht komplex programmiert sind, sondern sich ihre Komplexität aus der Summe von einfacheren Bauteilen ergibt. Dem zu Grunde liegt der einfachste Baustein, das sogenannte Perzeptron. Es handelt sich um ein einfaches vorwärtsgerichtetes Netz ohne innere Schicht, welches eine binäre Ausgabe (0 oder 1) liefert. Das Konvergenz-Theorem besagt, dass es die Fähigkeit besitzt in endlicher Zeit jede erlernbare Funktion anzutrainieren.[9] Ein Perzeptron eignet sich daher gut für die Mustererkennung, wobei jedes Perzeptron auf genau ein Merkmal antrainiert ist. Wie im späteren Fallbeispiel (siehe Kapitel 4) noch ersichtlich wird, werden viele simple Perzeptrons übereinandergeschichtet um gesamtheitlich ein künstliches neuronales Netzwerk mit einem sehr komplexen, genauen Filter- und Mustererkennungsmechanismus zu ergeben, ähnlich den Vorgängen im menschlichen Auge. Ein anderes faszinierendes Beispiel für die Funktionsweise von KI ist das Projekt NETtalk, welches 1986 von Sejnowski und Rosenberg durchgeführt wurde. Sie haben einem künstlichen neuronalen Netzwerk beigebracht Wörter und Buchstaben menschenähnlich auszusprechen. Zuerst wurde das System mit 1000 zufälligen Wörtern gefüttert, welche in einem Simulator 69 Stunden lang jegliche Aussprachevariationen durchspielten, bis das System eine Zielaussprachekorrektheit von 95% lieferte. [10] Anhand der folgenden Graphik kann das Konzept der KI nachvollzogen werden:

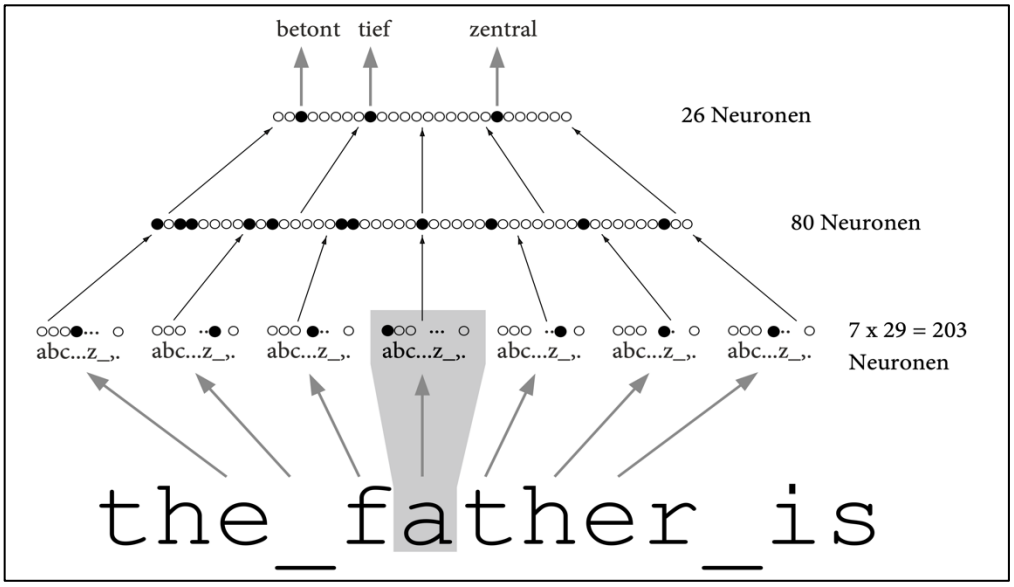


Abbildung 2: Das NETtalk-Netzwerk bildet einen Text auf seine Aussprachеattribute ab [11]

[8] vgl. Novustat, Künstliches neuronales Netz einfach erklärt - NOVUSTAT (13.August 2023)
[9] vgl. Lämmel U. / Cleve J. (2020), S. 203f
[10] vgl. Ertel W. (2021), S. 317
[11] Ertel W. (2021), S. 317

Nachdem das trainierte Netzwerk die Textdatei: „the_father_is“ erhält beginnt es sofort den Begriff in einzelne Buchstaben zu zerlegen, und mit allen Aussprachemöglichkeiten zu vergleichen. Nach einem komplexen Zusammenspiel der hidden layers (80 verdeckte Neurone) und Gewichtungen (dazu mehr im Kapitel 4.1.3) zwischen diesen Neuronen wird letztendlich die am wahrscheinlichsten korrekte Aussprachevariante ausgewählt und rekonstruiert.
Diese spannende Demonstration verdeutlicht die Fähigkeit von KI aus großen Datenmengen Muster zu erkennen und aus Erfahrungen interne Verbindungen zu ergänzen, um akkurate Ergebnisse zu erzielen.

Als nächstes werden die **Grundlagen des Gehirns** näher beleuchtet:
Die Großhirnrinde des menschlichen Gehirns ist verantwortlich für höhere kognitive Funktionen.[12] Es nimmt Sinneseindrücke (Bsp.: visuelle, haptische oder auditive Reize) auf, verarbeitet diese, und schickt Botschaften und Befehle (Bsp.: Muskelaktivierung, Ausweichen vor einem Hindernis, Musikerkennung, Hormonausschüttung) in alle Bereiche des Körpers zurück. Das Gehirn arbeitet mit elektrischen Signalen, die über Neuronen und Nervenzellen übertragen werden. Die Neuronen sind wiederum durch Synapsen verbunden, welche Neurotransmitter, chemische Botenstoffe, freisetzen. Diese Neurotransmitter sind mitverantwortlich für die Neuroplastizität des Gehirns. Darunter wird verstanden, dass die Verzweigungen des Gehirns sich laufend verändern, umvernetzten und verstärken, wenn etwas Neues geübt oder erlernt wird.[13] Diese einzigartige Anpassungsfähigkeit bildet die Grundlage für die menschliche Fähigkeit, sich ständig weiterzuentwickeln. Dies wird von der folgenden Abbildung vereinfacht veranschaulicht:

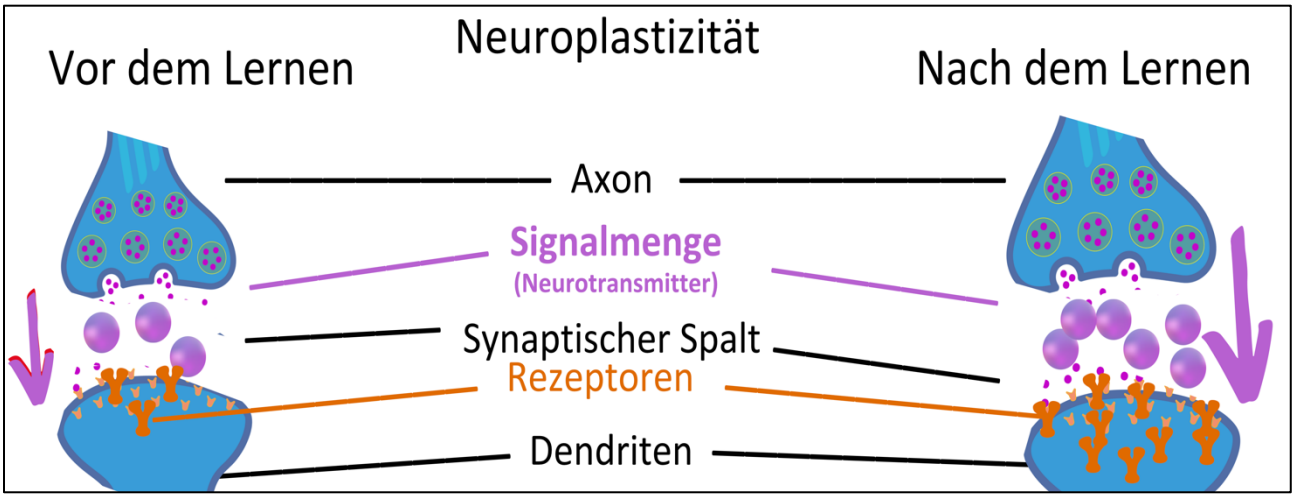


Abbildung 3: Visualisierung der Neuroplastizität[14]

[12] vgl. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG), Wie funktioniert das Gehirn? | Gesundheitsinformation.de (13. August 2023)
[13] vgl. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG), Wie funktioniert das Gehirn? | Gesundheitsinformation.de (13. August 2023)
[14] Eigene Darstellung in Anlehnung an Neusitzer H., Mein Ressourcencoach - Hartmut Neusitzer Vortrag Workshop & Coaching mit Zufriedenheitsgarantie - Glossar: Zürcher Ressourcen Modell, PSI-Theorie, Persönlichkeit (mein-ressourcencoach.de) 13. August 2023

Noch dazu lässt sich das Gehirn in unterschiedliche Hirnregionen unterteilen, die verschiedene Aufgaben behandeln. So ist die Amygdala für die Gefühlsverarbeitung zuständig, das Frontalhirn dient dem logischen Denken, und der Occipitallappen der visuellen Wahrnehmung. Diese Regionen stehen jedoch in ständigem Kontakt zueinander, leiten Informationen weiter und verarbeiten sie auf unterschiedliche Weise. Aus diesem Grund feuern in unserem Gehirn auch immer Neuronen.
Bei der visuellen Reizverarbeitung verläuft der Verarbeitungsprozess vereinfacht wie folgt: Retina -> Sehnerv -> primärer visueller Pfad -> Thalamus -> Hinterlappen/primärer visueller Kortex. Dieser langkettige Verarbeitungsprozess veranschaulicht die Komplexität und das Zusammenspiel der Gehirnregionen, das bei scheinbar einfachen Aufgaben, in unserem Kopf abläuft. Im näheren wird auf die spezifischen Gehirnprozesse in den folgenden Kapiteln eingegangen.

3 Neue Horizonte - Parallele zwischen Gehirnfunktionen und KI-Algorithmen

Das folgende Kapitel, geht näher auf die Analogien zwischen KI-Algorithmen und der neuronalen Funktionsweise im Gehirn ein. Dabei wird erörtert **auf welche Weise sich die Arbeitsweisen des menschlichen Gehirns in den Funktionsprinzipien von KI-Algorithmen widerspiegeln**.

Um die Einflüsse der neuronalen Prozesse im Gehirn auf KI-Systeme zu untersuchen, behandelt diese Arbeit verschiedenen Untergebiete, die klare Analogien in der Struktur und Funktionsweise aufweisen.
Es werden bewusst unterschiedliche Aspekte der KI analysiert, die anhand von Beispielen und Analogien zu ihrer ursprünglichen Inspiration im menschlichen Gehirn verglichen werden. Die Aufmerksamkeit liegt hierbei auf folgenden Hauptgebieten:

A. Struktur der Neuronen
B. Genetische Algorithmen
C. Convolutional Neural Network vs. Präfrontaler Kortex (Neuroplastizität)
D. Mathematische und statistische Methoden der KI
E. Bias

3.1 Struktur der Neuronen (Künstlich und Biologisch)

Bevor tiefer in die Unterthemen eingetaucht wird, werden kurz die grundlegenden Arbeits- und Funktionsweisen einander gegenübergestellt. Am deutlichsten sieht man diesen Kontrast bei dem Vergleich von menschlichen Neuronen und künstlichen Neuronen, die so viele Gemeinsamkeiten teilen, dass sogar viele Fachbegriffe übernommen wurden. Im folgenden Absatz wird ein vereinfachter Vergleich der Neuronen gezogen, um danach näher auf zwei Besonderheiten näher eingehen zu können.

Die neuronalen Dendriten, die in der Biologie als Kanäle für Eingangssignale dienen, finden ihre technologische Entsprechung in einem Eingangsvektor. Dieser Vektor wird mit Synapsengewichten multipliziert, eine Analogie zu dem Prozess, in dem der Mensch in seinen Synapsen chemische Botenstoffe und Neurotransmitter kombiniert, um ein kohärentes Signal zu ergeben.[15] Dieser Ausgangswert wird von dem künstlichen Axon an die Zelle weitergeleitet. Schon an diesem Punkt werden die Parallelen deutlich, die zeigen, wie menschliche neuronale Mechanismen in KI-Algorithmen reflektiert sind. Die nachstehende Abbildung illustriert diese Korrelation:

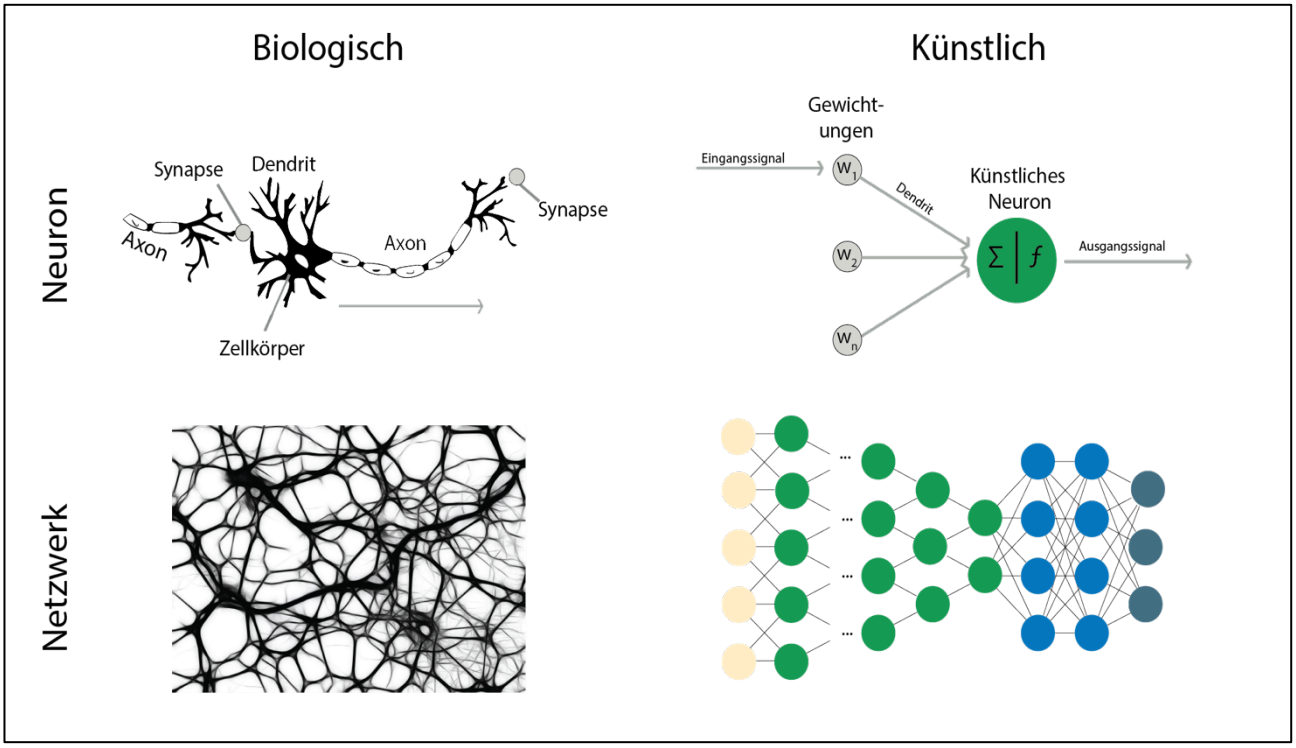


Abbildung 4 : Schematischer Vergleich von künstlichen und biologischen Neuronen[16]

[15] vgl. Steinwendner J. / Schwaiger R. (2020), S. 34
[16] Wäldchen J. / Mäder P., https://besjournals.onlinelibrary.wiley.com/doi/10.1111/2041-210X.13075, 20.09.2023

3.1.1 Aktivierungsfunktion vs. Schwellenwert bei der Signaltransduktion

Eine Vertiefung in die Thematik offenbart, dass künstliche Neuronen mehr als nur eine grobe Imitation der Strukturen und Terminologien unseres Gehirns sind. Sie berücksichtigen sogar komplexe biologische Feinheiten wie das Alles-oder-Nichts-Gesetz, und erfolgreich in einem technologischen Kontext um. Das **biologische Alles-oder-Nichts-Gesetz** „[..] bezeichnet das Phänomen, dass eine Reaktion auf einen Reiz entweder vollständig oder überhaupt nicht ausgelöst wird.“[17] Um eine Reaktion auszulösen muss ein kritisches Schwellenpotential überschritten werden. Dieses Prinzip ist besonders bedeutsam für Reize, die so schwach sind, dass sie keine Reaktion erfordern.

In der **künstlichen Intelligenz** finden wir ein Analogon zu diesem Gesetz in den **Aktivierungsfunktionen**. Interessanterweise existieren hier, sowohl mathematisch als auch technologisch, unterschiedliche Varianten dieser Aktivierungsschwellen, die den vielschichtigen Charakter der neuronalen Mechanismen des Menschen widerspiegeln. Die folgende Abbildung veranschaulicht unterschiedliche informatische Aktivierungsfunktionen, und hebt die unterschiedlichen Schwellenwerte und graphischen Verläufe hervor.

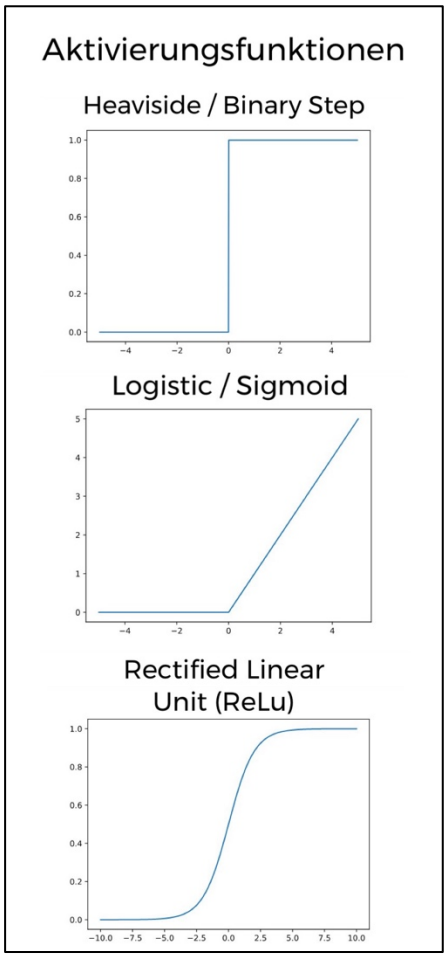


Abbildung 5: Aktivierungsfunktionen [19]

Die Heaviside Aktivierungsfunktion als Nachbau des Alles-oder-Nichts-Gesetzes ist ein anschauliches Beispiel, das verdeutlicht, wie sich die Arbeitsweisen des Gehirns in der KI widerspiegeln. Die ReLu- und Sigmoidaktivierungsfunktionen, die eher ein abgewandeltes „[..]Fast-Alles-oder-Fast-Nichts-Gesetz[..]“[18] nachbilden, zeigen exemplarisch, wie die KI natürliche Vorbilder oftmals abwandelt, optimiert und verfeinert, um sich auf die technologischen Ansprüche und Zielsetzungen des Modells anzupassen.

Die Abwandlung ermöglicht in diesem Fall, dass künstliche neuronale Netze, die mit einer Sigmoidaktivierungsfunktion ausgestattet sind, bei ihren Antworten feine Abstufungen machen können, und bei ihren Berechnungen Zwischenwerte annehmen können. Dies führt nicht nur zu einer enormen Steigerung der Leistung und Flexibilität der Modelle, sondern eröffnet auch eine neue Dimension, in der die Maschine die Fähigkeit hat, nicht nur Schwarz / Weiß oder 0 / 1 zu unterscheiden, sondern auch Graustufen zu erkennen und ohne Überschreitung eines harten Schwellwerts flexibel und adaptiv Antworten feinzustimmen.

---

[17] Biologie Seite, Alles-oder-nichts-Gesetz – biologie-seite.de, 02.09.2023
[18] Steinwendner J. / Schwaiger R. (2020), S. 35
[19] Eigene Darstellung (siehe Anhang 2)

3.2 Genetische Algorithmen

Genetische Algorithmen sind ein weiteres Beispiel, das die Natur als Inspirationsquelle für die Entwicklung in der KI nutzt. Sie folgen grundlegenden evolutionären Prinzipien wie "Survival of the Fittest," "natürliche Selektion," und "Kreuzung," was sie besonders attraktiv für Anwendungen in den Bereichen der Such- und Optimierungsprobleme sowie der Spiele- und Fahrzeugentwicklung macht.

Zunächst wird eine initiale Generation von Individuen, gemeinhin als Lernparameter bekannt, erzeugt. Jeder dieser Lernparameter stellt eine potenzielle Lösung des gegebenen Problems dar und wird durch eine Fitnessfunktion bewertet, um seine Leistung und Fehleranfälligkeit zu messen.

Darauf folgt eine Selektionsphase welche die besten Lernparameter auswählt und miteinander kreuzt, was dazu führt, dass das Programm von Generation zu Generation leistungsfähiger und fehlerresistenter wird. Zusätzlich wird eine kontrollierte Mutationsrate implementiert, um die Einführung neuer Eigenschaften und Lösungsvarianten zu ermöglichen.[20] Es lassen sich starke Parallelen zu den evolutionären Theorien von Herbert Spencer und Charles Darwin ziehen, insbesondere zu den Konzepten des "Survival of the Fittest," die durch Fitnessfunktionen identifiziert werden, und der natürlichen Selektion, welche nur die erfolgreichsten Codes oder Parameter in die nächste Generation überführt und sie miteinander kreuzt. Damit wird ein weiteres KI-Konzept deutlich, das in hohem Maße auf den Funktionsprinzipien des menschlichen Gehirns und der natürlichen Evolution basiert.

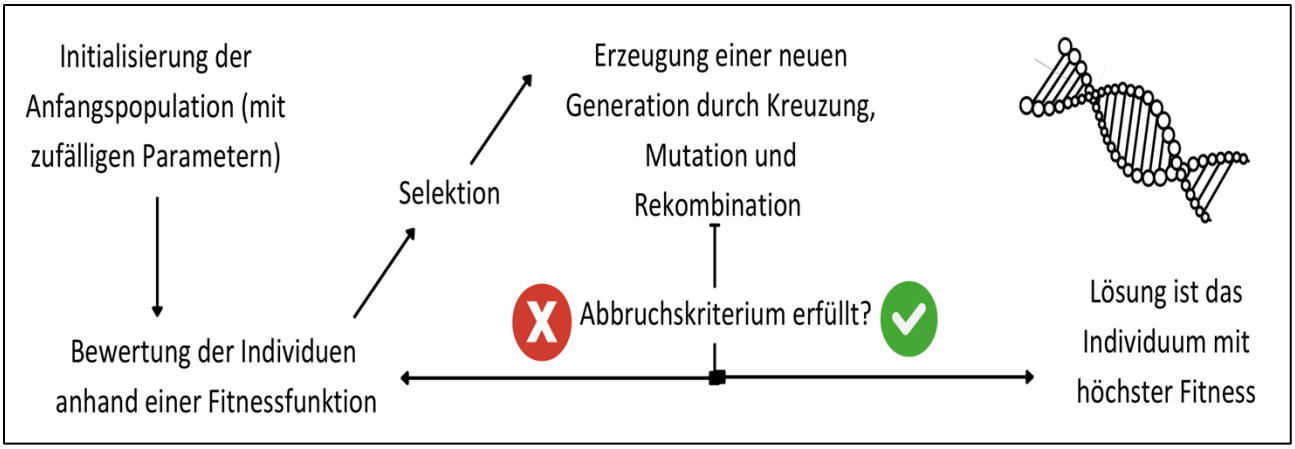


Abbildung 6: Schematische Darstellung genetischer Algorithmen [21]

[20] vgl. Steinwendner J. / Schwaiger R. (2020), S. 35
[21] Eigene Darstellung in Anlehnung an Stein. S, 5.3.4 Genetische Algorithmen (hpfsc.de), 1.1.2024

4 Experimentelle Exploration - Fallbeispiel eigenes KI-Modells

Zur Untersuchung von Korrelationsaspekten anhand eines konkreten Fallbeispiels wird im nachfolgenden Kapitel eine eigens entwickelte KI vorgestellt. Bei dem Code handelt es sich um ein Convolutional Neural Network (CNN). Dabei handelt es sich um ein „spezielles mehrschichtiges Feed-forward Netzwerk“[22]. Es besteht aus simplen Perzeptrons, die durch ihre vielen (mehrschichtig) Verbindungen eine sehr komplexe Filterbank bilden. Des Weiteren ist es vorwärtsgerichtet (Feed-forward), das heißt die Daten fließen durch mehrere versteckte Schichten, die jeweils eine simple Operation durchführen, bis sie die Ausgabeschicht erreichen und eine Klassifikation ausführen. Ausschlaggebend ist dabei, dass keine Rückkopplung oder Schleifen verwendet werden, wie bei vielen anderen KI-Modellen. CNNs werden hauptsächlich für maschinelle Klassifizierungsverfahren von Bild- oder Audiodateien gebraucht.[23] In meinem Beispiel dient es der Ziffernerkennung. Das Modell wurde mit einem Open-Source* Datensatz von „kaggle.com“ trainiert und in der Coding-Umgebung „Virtual Studio Code“ (VS Code) unter Verwendung von TensorFlow entwickelt.[24]

4.1 Convolutional Neural Network (CNN)

Dieses Convolutional Neural Network dient der Erkennung von handgeschrieben Ziffern.

4.1.1 Architektur

Das Programm wurde anhand des MNIST Datensatzes trainiert. Dieser verfügt über 70.000 Aufnahmen, die sich aus handgeschriebenen Ziffern zwischen 1 und 9 ergeben.[25] Ein Ausschnitt aus dem Datensatz ist in der Abbildung rechts zu erkennen.
In dem folgenden Absatz wird die Architektur eines CNN näher erläutert. Ein typisches CNN besteht aus einem Kodierungsblock und einen Prädiktionsblock.

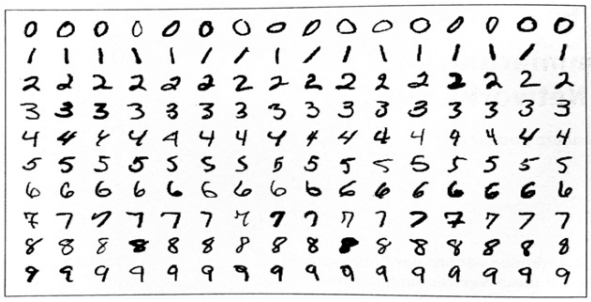


Abbildung 7: Ausschnitt aus dem MNIST Datensatz [26]

Der Kodierungsblock setzt sich aus einer Folge von Faltungs-, Aktivierungs-, und Poolingschichten zusammen welche Schlüsseleigenschaften des Bildes extrahieren und eine kodierte Repräsentation ergeben. Dies ist vergleichbar mit dem Lesen einer Fremdsprache, die man zuerst im Kopf in die eigene Muttersprache übersetzt werden muss, damit der Text begriffen werden kann. Auf ähnliche

[22] Steinwendner J. / Schwaiger R. (2020), S. 197
[23] vgl. Papp et al. (2022), S. 246ff
[24] vgl. Lämmel U. / Cleve J. (2020), S. 321
[25] vgl. Microsoft, https://learn.microsoft.com/de-de/archive/msdn-magazine/2014/june/test-run-working-with-the-mnist-image-recognition-data-set#der-mnist-datensatz, 22.08.2023
[26] Steinwendner J. / Schwaiger R. (2020), S. 222

Weise muss die KI die eingegebene Bilddatei in ein Format von Kodierungsvektoren übersetzen, um den Inhalt und die Bedeutung zu begreifen und daraus Schlüsse ziehen zu können.
Der Prädiktionsblock ist hingegen meistens ein herkömmliches neuronales Netz, welches die kodierte Repräsentation des Bildes klassifiziert.[27]
Diese Klassifikation erfolgt nach dem Softmax-Prinzip, das heißt, dass alle Aktivierungen der Neuronen der Ausgabeschicht normiert werden und so prozentuelle Wahrscheinlichkeiten der Klassenzugehörigkeiten ergeben, die eine Prädiktion ermöglichen.[28] Vereinfacht kann man das mit einem „educated guess" gleichsetzen, den wir Menschen benutzen, um anhand unserer Erfahrung eine Schätzung abzugeben, indem wir uns für die Antwort entscheiden, die mehr Neuronen in unserem Gehirn aktiviert.

Das CNN besitzt drei Convolutional -Schichten, gefolgt von Flattening- und Dense-Schichten. „Convolution" bedeutet übersetzt „der mathematische Begriff der Konvolution oder Faltung"[29], und steht im Zusammenhang mit der Bildverarbeitung für Filterbanken einfacher Merkmale (Kontrast, Kanten, Lichtfrequenzen, Punkte, Ecken, etc.).[30] Die folgende Tabelle veranschaulicht die Funktion und den Zuständigkeitsbereich der einzelnen Schichten (eng. layers):

Schicht	**Funktion**	**Zuständigkeit**
Convolutional 2D Layer	Faltungsoperationen auf dem Eingangsbild	Extraktion von Schlüsselmerkmalen wie Kanten, Texturen und Formen
Max Pooling 2D Layer	Pooling-Operation, Erstellung einer Merkmalskarte	Reduziert Dimensionen, betont Schlüsselmerkmale, verringert die Berechnungslast und verhindert Überanpassung (overfitting)
Flattening Layer	Umwandlung der 2D Merkmalskarten in 1D Vektor	Vorbereitung der Daten in einem lesbaren Format für die nachfolgende voll vernetzte Schicht mit Softmax Ausgabe
Fully Connected Layer	Dense-Schicht mit ReLU-Aktivierungsfunktion / Softmax Aktivierungsfunktion	Transformiert die Daten + ermöglicht das Lernen komplexer Zusammenhänge zwischen den Merkmalen/ Gibt die Klassifikationsergebnisse aus
Dropout Layer	Deaktiviert während des Trainings zufällig Neuronen (Dropout) und verhindert so Überanpassung	Verbessert die Leistung des Modells gegenüber Rauschen (Verschobene / Verschwommene Eingangsbilder) und verhindert, dass das Modell zu sehr auf bestimmte Merkmale trainiert wird (sonst keine Generalisierungsfähigkeit)

Tabelle 1: Funktion und Zuständigkeit der CNN Layers[31]

[27] vgl. Steinwendner J. / Schwaiger R. (2020), S. 199
[28] vgl. Lämmel U. / Cleve J. (2020), S. 257
[29] Steinwendner J. / Schwaiger R. (2020), S. 199
[30] vgl. Papp et al. (2022), S. 246
[31] vgl. Steinwendner J. / Schwaiger R. (2020), S. 199

Anhand des Codes erkennt man, dass viele dieser Schichten mehrmals wiederholt werden. Die Schichten arbeiten ähnlich, arbeiten sich aber immer von dem letzten Ergebnis vorwärts und können zu unterschiedlichen Zeitpunkten andere, komplexere Merkmale erfassen und priorisieren.

Die folgende Abbildung, die unter Verwendung eines open-source Codes für die verschiedenen Schichten in Latex erstellt und adaptiert wurde[32], veranschaulicht, wie die KI die Eingangsdatei in immer kleinere Teile zerlegt, während sie unterschiedliche Filter anwendet. Diese Funktionsweise ist eine Besonderheit der CNNs, welche erlaubt die Rechenleistungsgrenzen der konventionellen Netze zu sprengen und auch räumlich verschobene Objekte problemlos zu erkennen.[33]

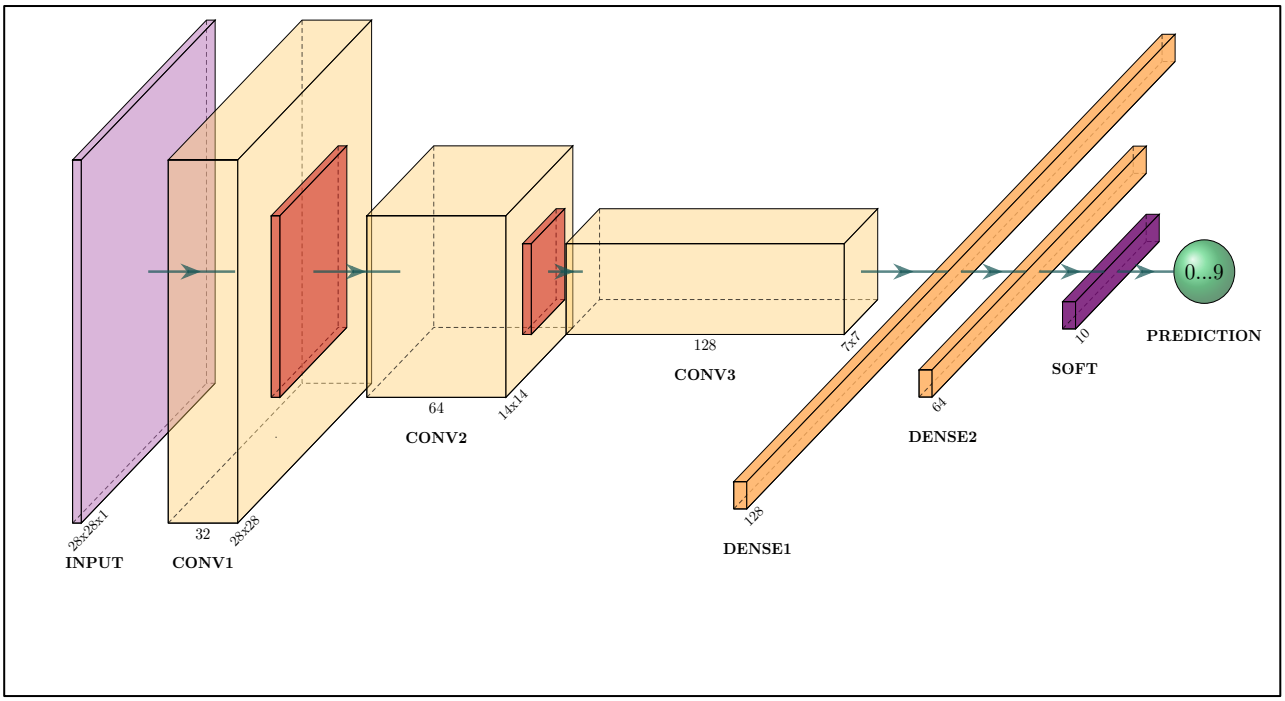


Abbildung 8: Struktur eines Convolutional Neural Networks[34]

Auf der folgenden Seite befindet sich der Hauptteil des Codes der CNN, mit Kommentaren (mit # über den jeweiligen Befehlszeilen angezeigt) die die Bedeutung der Python-Befehle erklären.

[32] vgl. GitHub, https://github.com/HarisIqbal88/PlotNeuralNet, 20.12.2023
[33] vgl. Lämmel U. / Cleve J. (2020), S. 252
[34] Eigene Darstellung (siehe Anhang 3)

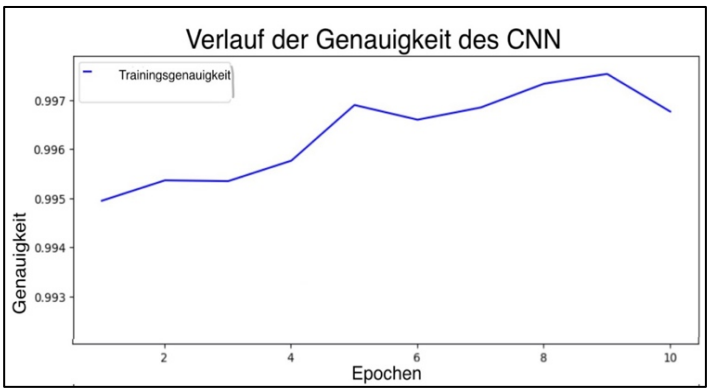
```
# Convolutional Neural Network (CNN) zur Ziffernerkennung
# Tensorflow Framework importieren
import tensorflow as tf
# MNIST Datensatz importieren (handgeschriebene Ziffern)
from tensorflow.keras.datasets import mnist
# Import einer Sequential Klasse um ein sequentielles Netzwerk zu erstellen
from tensorflow.keras.models import Sequential
# Import von verschiedenen Schichten
from tensorflow.keras.layers import Conv2D, MaxPooling2D, Flatten, Dense, Dropout
# Import der Matplotlib Bibliothek
import matplotlib.pyplot as plt

__author__ = „Larissa Arthofer“

# Laden des MNIST Datensatz, Aufteilen in Übungs- und Testdaten
(train_images, train_labels), (test_images, test_labels) = mnist.load_data()
# Konvertierung der Bilder in gewünschtes Format
train_images = train_images.reshape((60000, 28, 28, 1)).astype('float32') / 255
# (28x28 Pixel mit einer Tiefe 1), Skalierung auf Wertebereich [0, 1]
test_images = test_images.reshape((10000, 28, 28, 1)).astype('float32') / 255

# Umwandlung der Labels in ein One-Hot Encoding
train_labels = tf.keras.utils.to_categorical(train_labels)
# Aus numerischen Klassenlabels (0 bis 9) werden binäre Vektoren erstellt
test_labels = tf.keras.utils.to_categorical(test_labels)

# CNN-Modell erstellen
model = Sequential()
# 2D-Faltungsschicht, extrahiert Merkmale aus den Bilders
model.add(Conv2D(32, (3, 3), activation='relu', input_shape=(28, 28, 1)))
# Pooling-Schicht, reduziert die Dimensionen der Merkmalskarten
model.add(MaxPooling2D((2, 2)))
# Flatten-Schicht, wandelt Merkmalskarten in einen flachen Vektor um
model.add(Conv2D(64, (3, 3), activation='relu'))
model.add(MaxPooling2D((2, 2)))
model.add(Conv2D(128, (3, 3), activation='relu'))
model.add(Flatten())
# Dense-Schicht, "verdichtet" und verbindet die Neuronen miteinander
model.add(Dense(128, activation='relu'))
# Dropout-Schicht inkl. Dropout-Regulierungen verhindert Überanpassungen
model.add(Dropout(0.5))
# Letzte Dense Schicht, hat 10 Neuronen für die Ausgabe der Klassifizierung
model.add(Dense(64, activation='relu'))
model.add(Dense(10, activation='softmax'))

# Modell kompilieren - Optimierer, Verlustfunktion und Genauigkeitsmetrik festlegen
model.compile(optimizer='adam',loss='categorical_crossentropy',metrics=['accuracy'])

# Modell mit Trainingsdaten trainieren, 10 Epochen Laufzeit, Batch-Größe 128
model.fit(train_images,train_labels,epochs=10,batch_size=128, validation_data=(test_images, test_labels))

# Überwachung der Leistung durch Validierungsdaten
```

Code 1: Convolutional Neural Network zur Ziffernerkennung

4.1.2 Trainingsverlauf

Nach der Entwicklung des Codes begann das eigentliche Training. Dazu wurde der MNIST-Datensatz zuallererst in 60.000 Trainingsdaten und 10.000 Testdaten geteilt. Das Training erstreckte sich über 10 Epochen mit einer Batch-Größe von 128 und zeigt eine Analogie zur Evolution: Generationen von Lebewesen passen sich schrittweise an ihre Umgebungen und Verhältnisse an, so wie in dem Simulationsprozess die Software sich über mehrere Epochen immer mehr adaptiert, um sein gewünschtes Ziel zu erreichen.

Während des Trainings erzielte das Modell auf den Validierungsdaten eine Genauigkeit (eng. accuracy) von 99,4%. Die folgenden Abbildungen veranschaulichen die Leistungssteigerung der KI im Laufe der verschiedenen Trainingsepochen:




```
33 # programmiert von Larissa Arthofer
```

PROBLEMS 3 OUTPUT DEBUG CONSOLE TERMINAL PORTS 7

```
469/469 [==============================] - 33s 65ms/step - loss: 0.2996 - accuracy: 0.9065 - val_accuracy: 0.9831
Epoch 2/10
469/469 [==============================] - 29s 62ms/step - loss: 0.0796 - accuracy: 0.9766 - val_accuracy: 0.9871
Epoch 3/10
469/469 [==============================] - 30s 63ms/step - loss: 0.0555 - accuracy: 0.9839 - val_accuracy: 0.9905
Epoch 4/10
469/469 [==============================] - 29s 62ms/step - loss: 0.0438 - accuracy: 0.9874 - val_accuracy: 0.9905
Epoch 5/10
469/469 [==============================] - 29s 62ms/step - loss: 0.0362 - accuracy: 0.9900 - val_accuracy: 0.9896
Epoch 6/10
469/469 [==============================] - 29s 62ms/step - loss: 0.0300 - accuracy: 0.9913 - val_accuracy: 0.9917
Epoch 7/10
469/469 [==============================] - 30s 64ms/step - loss: 0.0266 - accuracy: 0.9918 - val_accuracy: 0.9917
Epoch 8/10
469/469 [==============================] - 29s 62ms/step - loss: 0.0230 - accuracy: 0.9931 - val_accuracy: 0.9915
Epoch 9/10
469/469 [==============================] - 29s 62ms/step - loss: 0.0202 - accuracy: 0.9940 - val_accuracy: 0.9933
Epoch 10/10
469/469 [==============================] - 30s 64ms/step - loss: 0.0187 - accuracy: 0.9944 - val_accuracy: 0.9932
```

Code 2: Trainingsverlauf (Ausgabe des CNN)[35]

In der dargestellten Abbildung des Trainingsverlaufs fällt auf, dass bereits am Ende der ersten Epoche eine bemerkenswert hohe Genauigkeit erzielt wurde. Des Weiteren sind signifikante Sprünge in den Werten für Genauigkeit und Verlustfunktion zwischen den einzelnen Trainingsepochen zu beobachten. Es zeigt sich, dass die Verbesserung der Modellgenauigkeit kein linearer Prozess ist, was auf die Komplexität des Optimierungsprozesses hindeutet.

Ein besonderes Augenmerk verdient der drastische Sprung in der Verlustfunktion zwischen der ersten und der zweiten Epoche, von einem Wert von 0.29996 auf 0.0796. Dies weist auf eine

[35] Eigene Darstellung

effiziente Anpassung der Gewichtungen im Netzwerk in der initialen Phase des Trainings hin, wobei der Algorithmus bereits in der Lage war, essentielle Merkmale der Trainingsdaten effektiv zu erfassen und für die Klassifikation zu nutzen.

Die Sprünge der KI-Leistung zwischen den einzelnen Epochen liegen an dem Finetuning der Parameter, also dem Ändern und Aktualisieren der Gewichtungen am Ende von jeder Epoche. Um diese Lernkurve näher zu untersuchen habe ich den Code ein zweites Mal ausgeführt und eine Graph-Funktion eingebaut, die die Änderung der Genauigkeit zwischen den Epochen als Diagramm veranschaulichen soll. Bemerkenswert war hierbei, dass die KI beim zweiten Trainingslauf ihre Genauigkeit noch weiter verbesserte und nach 10 weiteren Epochen eine Genauigkeit von 99,7% erreicht hat.

Auffallend sind aber auch die periodischen Einbrüche der Genauigkeitskurve. Dies ist dem Phänomen der Überanpassung (Overfitting) zuzuschreiben. In dem Fall werden „statistisch gesehen [..] zur Spezifizierung eines Modells zu viele erklärende Variablen eingesetzt.“[36] Durch die Integration von übermäßig vielen erklärenden Variablen wird die Generalisierungsfähigkeit des Modells beeinträchtigt und die Leistung vermindert. Eine analoge Situation erlebe ich als ambitionierte Tennisspielerin. Spieler, die ausschließlich mit konventionellen, einfachen Schlägen in Trainerstunden trainieren, zeigen zwar eine nahezu fehlerfreie Ausführung dieser Schläge, allerdings mangelt es ihnen an Anpassungsfähigkeit gegenüber neuen Gegnern und ungewohnten Spielsituationen. Dies spiegelt die Kernproblematik der Überanpassung in KI-Modellen wider: Sie optimieren ihre Leistung hinsichtlich der Trainingsdaten, verlieren aber die Fähigkeit zur Generalisierung und Anpassung. Ein weiteres, alltägliches Beispiel für das Phänomen der Überanpassung findet sich im schulischen Kontext. Besonders auffällig wird dies in Fächern, in denen ein tiefgehendes Verständnis der Theorien erforderlich ist. Schüler, die sich darauf konzentrieren, spezifische Aufgabenstellungen auswendig zu lernen, stoßen an ihre Grenzen, wenn sie mit neuen, unbekannten Problemen konfrontiert werden. Diese Limitierung ist vergleichbar mit KI-Modellen, die aufgrund von Überanpassung an Trainingsdaten, ihre Fähigkeit zur Problemlösung in unbekannten Szenarien verlieren.[37]

In Summe offenbart dies erneut eine Parallele zwischen den Funktionsweisen der KI und des menschlichen Gehirns, beide stoßen auf ähnliche Herausforderungen.

---

[36] Thamm A., Overfitting - [at] Data Science & KI Glossar (alexanderthamm.com), 03.09.2023
[37] vgl. Hinterleitner, Isabella: Interview, 24.08.2023, siehe Anhang 7

4.1.3 Erkenntnisse: Neuroplastizität bei der Ziffernerkennung

In diesem Kapitel liegt der Fokus auf den zentralen Analogien, die durch das vorangegangene Fallbeispiel illustriert werden. Insbesondere wird die Simulation der Neuroplastizität des menschlichen Gehirns durch Convolutional Neural Networks (CNNs) erörtert.

Diese Erkenntnisse fungieren als solide Basis für die vertiefte Exploration, wie neurowissenschaftliche Forschung zur Optimierung von KI-Technologien in der Mustererkennung beiträgt. Neuroplastizität und die neuronale Konnektivität ermöglichen Menschen sich kontinuierlich anzupassen und zu lernen. Dies geschieht durch die Umstrukturierung neuronaler Verbindungen, um neue Informationen und Fähigkeiten zu integrieren. Einige Experten haben diese Analogie ebenfalls als eine der wichtigsten Prinzipien der KI identifiziert.

„Das wichtigste Prinzip: Neuronen sammeln Input aus den unterschiedlichen Quellen, integrieren diese Signale und leiten ein neu generiertes Signal an nachfolgende Neuronen weiter.
Das ist das einfach Prinzip, das die Basis der von uns verwendeten [CNN] darstellt und das wir von der Natur lernen durften."[38]

Im letzten Beispiel wurde aufgeschlüsselt, wie das Convolutional Neural Network aufgebaut ist, und aufgezeigt, dass CNNs während des Trainings lernen komplexe Gewichtungen und Verbindungen zwischen künstlichen Neuronen aufzubauen, um die Muster und Ziffern zu erkennen. Die interessante Parallele zur Neuroplastizität, ist einerseits die Fähigkeit diese gelernten Gewichtungen und Verbindung kontinuierlich anzupassen und andererseits, dass sie dieses Wissen für völlig neue Aufgaben wiederverwenden können. Diese Kapazität nennt sich Transfer Learning, weil sie ihr Wissen auf neue Problemstellungen „transferieren" können. Sie ähnelt in gewisser Weise der Neuroplastizität des menschlichen Gehirns, bei der bereits existierende neuronale Verbindungen für neue Aufgaben umgestaltet werden. Mein CNN, das für die Ziffernerkennung trainiert wurde, könnte beispielsweise die erworbenen Fähigkeiten nutzen um eine neue Aufgabe, zum Beispiel das Erkennen von Buchstaben zu erlernen, ohne von Grund auf neu trainiert werden zu müssen.

Die Stärkung von neuronalen Verbindungen durch wiederholte Aktivierung und die Degeneration dieser durch längerfristige Inaktivität, lassen sich auch bei KI-Modellen beobachten:

1) Gewichte, die während des Trainings kaum aktiviert werden, werden entfernt / umverteilt.

In der Praxis werden diese Verbindungen nicht gelöscht, sondern nur zufällig umverteilt, da eine totale Entfernung zur Verdünnung der Netztopologie und somit einen Verlust an Speicherplatz führen würde. Durch die Umverteilung kann sich die Verbindung bei den tatsächlich relevanten Gebieten einfügen und so die Leistung des Modells optimieren.

2) Eingabemuster ergeben unterschiedliche Regionen, die eine unterschiedlich starke Relevanz für den Erkennungsprozess haben. Diese Regionen werden in unterschiedliche rezeptive Felder geteilt, analog zu den verschiedenen Hirnregionen.

---

[38] Steinwendner J. / Schwaiger R. (2020), S. 263

Während wichtigere rezeptive Felder optimiert werden, indem sie mehr Verbindungen aufbauen, können irrelevante Eingabe-Neuronen schlussendlich gar keine Verbindungen mehr besitzen, weil diese Information für die Ausgabe unwichtig scheint.[39]

Die folgende Abbildung, die mit adaptiertem Code von Stack Exchange[40] in Latex erstellt wurde, veranschaulicht dieses Verfahren:

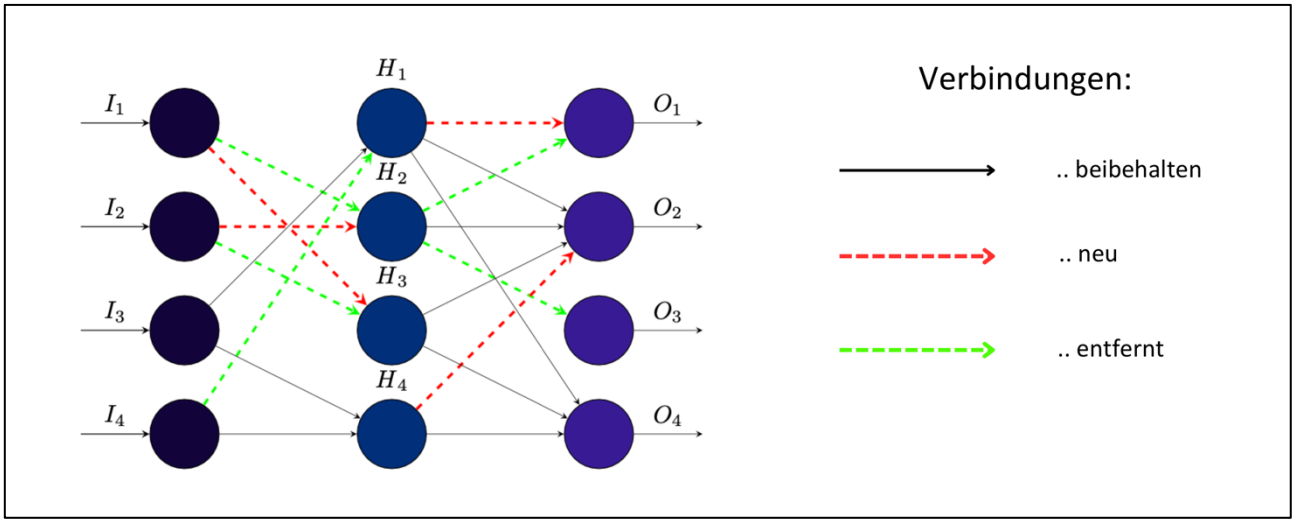


Abbildung 9: Umverteilung von Neuronenverbindungen[41]

Die maschinelle Nachahmung der Neuroplastizität, inspiriert von Erkenntnissen aus der Hirnforschung, tragen erheblich zur Struktur und Optimierung von KI-Technologien zur Mustererkennung bei.

Im Hinblick auf die Frage, wie Erkenntnisse aus der Hirnforschung zur Struktur und Optimierung von KI-Technologien zur Mustererkennung beitragen, wird deutlich, dass die Inspiration aus der Neuroplastizität nicht nur die Leistungsfähigkeit von KI-Systemen steigert, sondern auch die Ressourceneffizienz und Rechenleistung erhöht. KI-Modelle können aufgrund dieser adaptiven Fähigkeiten mit weniger Daten auskommen und dennoch effektive Ergebnisse erzielen. Dies eröffnet spannende neue Dimensionen für die Entwicklung von KI-Systemen, die in der Lage sind, sich kontinuierlich anzupassen und zu lernen. Darüber hinaus tragen ähnliche Erkenntnisse aus der Hirnforschung bezüglich der neuronalen Konnektivität auch dazu bei, KI-Systeme robuster gegenüber Änderungen und Störungen zu machen. So wie das menschliche Gehirn in der Lage ist, sich an veränderte Bedingungen anzupassen, sollen KI-Systeme flexibler und widerstandsfähiger gegenüber unvorhergesehenen und neuen Umständen werden.

[39] vgl. Lämmel U. / Cleve J. (2020), S. 244
[40] vgl. Stack Exchange, https://tex.stackexchange.com/questions/153957/drawing-neural-network-with-tikz, 1.1.2024
[41] Eigene Darstellung (siehe Anhang 4)

4.2 Vergleich mit dem präfrontalen Kortex

Um weitere Korrelationen zwischen den Prozessen im Gehirn und in der CNN aufzudecken, werden in diesem Kapitel die Arbeitsschritte bei der visuellen Verarbeitung verglichen.

4.2.1 Arten der Signalverarbeitung (Sequenziell oder Parallel)

Im Allgemeinen ist es bekannt, dass die KI uns bei der Muster- und Objektserkennung noch weit nachhingt. Dies ist größtenteils zurückzuführen auf die 100-Schritt Regel, welche in der Neuroinformatik dazu dient, die Leistungsfähigkeit des Gehirns zu demonstrieren.[42]

„Ein Mensch kann einen ihm bekannten Gegenstand oder eine bekannte Person innerhalb von 0,1 s erkennen. Dabei sind [..] maximal 100 sequentielle Verarbeitungsschritte im Gehirn des Menschen nötig.“[43]

Das wird einerseits durch die extrem schnelle Reaktionszeit der Neuronen im Millisekunden-Bereich (10^{-3}), sowie durch die parallele Verarbeitung der Informationen ermöglicht.[44] Die Objekterkennung findet also insgesamt in weniger als 100 sequentiellen Schritten statt. Jedoch kann die Anzahl der gesamten Verarbeitungsschritte die parallel erfolgt, weit über diese Zahl hinausschießen.[45] Die folgende Abbildung demonstriert diese unterschiedlichen Verarbeitungsabläufe und klärt auf, weshalb die Technologie in diesem Aspekt immer noch nicht mit unserem Gehirn mithalten kann.

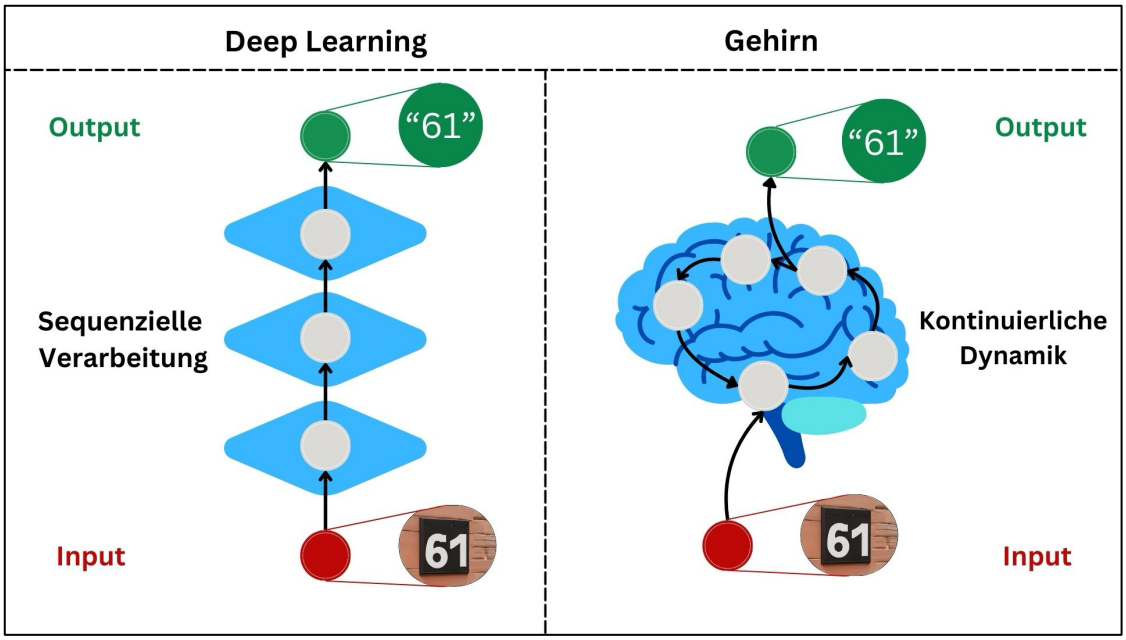


Abbildung 10: Verarbeitungsmethoden: Sequenziell vs. Parallel[46]

[42] vgl. Biologie Seite, 100-Schritt-Regel – biologie-seite.de, 03.09.2023
[43] Biologie Seite, 100-Schritt-Regel – biologie-seite.de, 03.09.2023
[44] vgl. Wikipedia, https://de.wikipedia.org/wiki/Künstliches_Neuron#Biologische_Motivation, 03.09.2023
[45] vgl. Biologie Seite, 100-Schritt-Regel – biologie-seite.de, 03.09.2023
[46] Eigene Darstellung in Anlehnung an Helmut L., https://www.golem.de/news/kuenstliche-intelligenz-wie-sich-deep-learning-vom-gehirn-unterscheidet-2202-162231-4.html, 04.09.2023

Nachdem der grundlegende Unterschied in der visuellen Informationsverarbeitung aufgezeigt wurde, widmet sich die Arbeit wieder den Gemeinsamkeiten in den sequenziellen Verarbeitungsschritten, die sowohl das CNN, als auch unseren präfrontalen Kortex, der zuständig für die Bildverarbeitung im Gehirn ist, betreffen. Das menschliche visuelle System ist ein äußerst komplexes System, welches eine Vielzahl von Strukturen umfasst. Die Retina wandelt visuelle Reize in neuronale Signale um, welche über den Thalamus zu dem primären visuellen Kortex, den sekundären visuellen Kortex Bereichen bis hin zu dem Assoziationskortex weitergeleitet werden, wo sie verarbeitet und interpretiert werden.[47]
Der Einfachheit halber wird jedoch nur auf die Strukturen eingegangen die Ähnlichkeiten zu dem CNN aufweisen. Das sind unter anderem die Vorverarbeitung der Eingangsdaten, die Hierarchie und Abstraktionsfähigkeit der Mustererkennung, sowie die Auswahl des Endergebnisses.

4.2.2 Vorverarbeitung der Eingangsdaten

Zu Beginn widmet sich diese Arbeit mit der Vorverarbeitung von Daten in dem CNN und der anfänglichen Informationsverarbeitung im Gehirn. Dafür werden die Vorverarbeitungsschichten in beiden Systemen betrachtet, um einen umfassenden Vergleich zu erstellen. Bei der CNN sind zwei Schichten besonders ausschlaggebend für die Vorverarbeitung: die Faltungsschichten (eng. Convolutional Layers) und die Pooling-Schichten (Max Pooling Layers). Diese werden beide mit ihrem Gegenstück im Gehirn verglichen.

Die Faltungsschicht der CNN ist analog zum visuellen Kortex im Gehirn. Beide sind zuständig für einfache Merkmale wie Kanten, Texturen und Formen aus den Rohdaten (Bilddatei oder visueller Reiz) zu extrahieren. In der CNN funktioniert diese Merkmalsextraktion durch Filterbanken, bestehend aus einer Stapelung vieler Kernels, die jeweils ein simples Merkmal suchen.[48] Auf ähnliche Art und Wiese läuft dieser Prozess im Gehirn ab, wo verschiedene rezeptive Bereiche des visuellen Kortex auf verschiedene visuelle Merkmale spezialisiert sind.

Die Pooling-Schichten der CNN reduzieren die Dimensionalität des Eingangsbildes, ähnlich wie im visuellen Kortex im Gehirn durch Pooling- und Downsamplingprozesse die räumliche Auflösung reduziert. Downsampling geschieht zum Bsp. im Sehnerv, wo die Informationen von Millionen von Rezeptoren auf knappe 1,2 Millionen Sehfasern reduziert werden.[49] Pooling ähnelt sich zwischen CNN und Gehirn sehr stark, im CNN werden die stärksten Schlüsselmerkmale herausgehoben und gespeichert, während im Gehirn die Informationen betont, werden die für Wahrnehmung und Verarbeitung von Bedeutung sind.[50] Diese Phänomene sind bedeutsam um visuelle Daten zu komprimieren ohne bedeutsamen Schlüsselmerkmale zu verlieren und um die Informationsverarbeitungseffizienz zu optimieren.

---

[47] vgl. Breiner. T (2018), S.122
[48] vgl. Steinwendner J. / Schwaiger R. (2020), S. 200f
[49] vgl. Wikipedia, https://de.wikipedia.org/wiki/Sehnerv, 03.09.2023
[50] vgl. Breiner. T (2018), S.122ff

4.2.3 Hierarchie und Abstraktion der Mustererkennung

In diesem Kapitel setzt sich die Arbeit näher mit dem hierarchischen Aufbau der verschiedenen Merkmalsdetektoren auseinander und geht auf die Abstraktionsfähigkeit bei der Merkmalsextraktion ein.

Die visuelle Informationsverarbeitung im Gehirn erfolgt hierarchisch und durch die verschiedenen Zelltypen in der Retina (Augennetzhaut). Die Zapfenzellen sind spezialisierte Rezeptoren, die für die Wahrnehmung von Kontrasten, Farben und feinen Details verantwortlich sind. Sie sind besonders in unserem zentralen Gesichtsfeld konzentriert und ermöglichen das Erkennen von Details, beispielsweise bei der Identifizierung von Ziffern. Sie funktionieren im Zusammenspiel mit den Stäbchenzellen, die lichtempfindlichen Rezeptoren die zuständig für das Erfassen von Helligkeitswerten und starken Lichtkontrasten sind. Sie sind in der peripheren Region der Retina verteilt und unterstützen die Wahrnehmung in dunklen Situationen sowie die Erfassung von Bewegungen. Die erfassten Merkmale und visuellen Informationen werden dann über den Sehnerv und den Thalamus, zum primären visuellen Kortex geleitet. Dort findet die Abstraktionsfähigkeit und das Erfassen unterschiedlicher Aspekte der visuellen Information statt.[52]

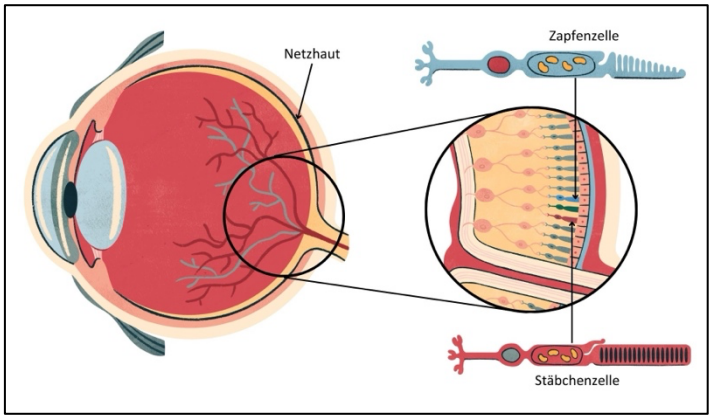


Abbildung 11: Anatomie der menschlichen Retina[51]

In dem CNN erfolgt eine ähnliche Hierarchie, obwohl sie auf künstlichen neuronalen Netzwerken basiert. Zuerst werden die niedrigen Schichten (Convolutional Layers) aktiviert. Diese Schichten sind analog zu den Zapfenzellen im menschlichen Auge. Sie dienen der Erkennung einfacher Merkmale wie Kanten, Textur und einfache Formen. In Bezug auf die Ziffernerkennung entsprechen sie dem ersten Schritt, bei dem grundlegende visuelle Merkmale erfasst werden.
Als nächstes kommen die höheren Schichten (Deep Layers) zum Einsatz. Diese Schichten entsprechen den komplexeren Verarbeitungsbereichen im menschlichen Gehirn. Hier werden in meinem CNN abstraktere Merkmale und komplexere Muster erkannt, die für die Ziffernerkennung von entscheidender Bedeutung sind.

Diese Hierarchie und die Fähigkeit zur Abstraktion der Merkmale ermöglichen es sowohl dem Gehirn als auch meinem CNN, allmählich von einfachen Merkmalen zu komplexeren Informationen überzugehen und somit eine effektive Mustererkennung und Ziffernerkennung zu gewährleisten. Dieser Vergleich verdeutlicht, wie KI-Modelle die Funktionsweise des menschlichen Gehirns in der Mustererkennung nachahmen und weiterentwickeln können, indem sie verschiedene Ebenen der Informationsverarbeitung nutzen, ähnlich den unterschiedlichen Rezeptoren im menschlichen Auge.

---

[51] Eigene Darstellung in Anlehnung an Heitling G., https://www.allaboutvision.com/de/augenanatomie/netzhaut/, 1.1.2024
[52] vgl. Fraunberger A., https://www.jungeroemer.net/blog/cyberspace-design-und-neurowissenschaften/, 10.09.2023

4.2.4 Klassifikationsresultat

Um diesen Abschnitt abzuschließen, wird ein Fokus auf den letzten Schritt in der Informationsverarbeitungskette gelegt, in dem das endgültige Klassifikationsergebnis berechnet oder getroffen wird. Im menschlichen Gehirn verläuft die Objekterkennung in drei Schritten: „Merkmale erkennen, Merkmale zu einem Objekt zusammensetzten und Objekt durch Rückgriff auf das Gedächtnis identifizieren.“[53] In Bezug auf die Identifikation existieren unterschiedliche Theorien, wobei in dieser Betrachtung die Schablonentheorie im Mittelpunkt steht, da sie die größten Übereinstimmungen mit dem CNN aufweist. Die Schablonentheorie geht davon aus, dass wir bereits Prototypen im Gedächtnis haben, die repräsentative Muster für bekannte Objekte sind. Wenn wir ein Objekt betrachten, wird das auf der Netzhaut hinterlassene Muster wie eine Schablone auf diese bekannten Prototypen angewendet. Das betrachtete Objekt wird als derjenige Prototyp identifiziert, dessen Muster am besten übereinstimmt. Wichtig ist, dass die Übereinstimmung nicht 100% sein muss, sondern eine ausreichende Übereinstimmung genügt.[54]
Wenn wir die Gültigkeit dieser Theorie annehmen, dann gibt es tatsächlich nur geringfügige Unterschiede zu der Klassifikationsweise einer KI, oder von meines CNNs. Vereinfacht ausgedrückt berechnet das CNN im letzten Schritt die prozentuelle Wahrscheinlichkeit für jede Kategorie, indem es die zuvor erfassten Merkmale mit einem Prototypmuster aus dem Datensatz vergleicht. Schließlich wählt es das Ergebnis mit der höchsten Wahrscheinlichkeit aus, also das, bei dem das Muster am meisten ähnelt.
Dies zeigt erneut ein Beispiel dafür, wie KI-Modelle die Funktionsweise des menschlichen Gehirns nachahmen und dabei dazu beitragen, die Grundlagen der Mustererkennung besser zu verstehen und weiterzuentwickeln.

4.3 Quantifizierung des Intellekts - Die statistische Anatomie der KI

Obwohl der Fokus dieser Arbeit bis jetzt primär auf den Parallelen zwischen KI und dem menschlichen Gehirn lag, widmet sich dieses Kapitel dem primären Herzstück der KI: dem mathematischen und statistischen Kern. Im Gegensatz zum menschlichen Intellekt, der auf biochemischen Prozessen und neuronalen Netzwerken basiert, operiert die KI auf einer grundlegend anderen Basis: einer rechnerischen.

Zwar nimmt sie in vielen Fällen Methoden und Erkenntnisse aus der Hirnforschung als Hilfsmittel in Anspruch, doch ihre Funktionsweise bleibt tief in der Mathematik und Statistik verwurzelt. Dieses Kapitel bietet einen selektiven Überblick über verschiedene rechnerische und statistische Methoden, die in der KI-Technologie zum Einsatz kommen, und rundet damit das Thema aus einer technologischen Perspektive ab. Besonderer Fokus liegt dabei auf die grundlegendste statistische Methode, die binäre logistische Regression, sowie einen der rechnerischen Ansätze, der verantwortlich für die gesteigerte Effizienz und Leistung der neuesten ChatGPT Modelle ist, nämlich die Vektormodellierung.

---

[53] Study Smarter, Objekterkennung: Verfahren & Funktionsweise | StudySmarter, 03.09.2023
[54] vgl. Study Smarter, Objekterkennung: Verfahren & Funktionsweise | StudySmarter, 03.09.2023

4.3.1 Binäre logistische Regression

Die binäre logistische Regression ist eine der weit verbreiteten und grundlegenden Methoden in der statistischen Modellierung und dient als wichtiger Baustein in der rechnerischen Funktionsweise der KI, insbesondere bei der Klassifizierung und Vorhersage binärer Ereignisse.

Die logistische Regression ist in der Künstlichen Intelligenz eine Schlüsselmethodik, die als Erweiterung der bivariaten linearen Regression mit der zusätzlichen Fähigkeit der Labelerstellung und Kategorienzuweisung entwickelt wurde.

In der bivariaten linearen Regression wird ein numerischer Ausgabewert erzeugt, der die kombinierten Effekte von zwei Variablen berücksichtigt. Diese Kombination stützt sich auf gewichtete Eingaben und eine bestimmte Fehlerquote, was die Methode sowohl leicht interpretierbar als auch effizient macht. Sie wird häufig in der Vorhersage von Verbrauchernachfrage, Kinoeinnahmen und Immobilienpreisen angewendet. Die Methode ähnelt dem Zeichnen einer Trendlinie, die auf der Analyse mehrerer simultaner Gleichungen basiert und die Verbindung zwischen verschiedenen Variablen darstellt.[55]

Die grundlegende Gleichung der logistischen Regression lautet:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 \times X_1 + \beta_2 \times X_2 + \cdots + \beta_k \times X_k$$

Hierbei ist p die Wahrscheinlichkeit eines positiven Ausgangs und $X_1, X_2, \ldots . X_k$ sind die unabhängigen Variablen. Die Koeffizienten $\beta_0, \beta_1, \ldots \beta_k$ sind die zu schätzenden Parameter. Die Wahrscheinlichkeit p lässt sich ausdrücken als:

$$p = \frac{1}{1 + e^{-(\beta 0 + \beta 1 \times X1 + \beta 2 \times X2 + \cdots + \beta k \times Xk)}}$$

Diese Modifikation ist notwendig, um die Einschränkungen der Linearen Regression zu überwinden und um die Labelerstellung und Kategorisierung von Daten zu ermöglichen. Dies ist nützlich bei der Erkennung von Zahlen oder der Differenzierung zwischen authentischen Nachrichten und fake News, und wird als binäre logistische Regression bezeichnet. Diese Methode dient der Vorhersage der Wahrscheinlichkeit für das Auftreten eines binären Ereignisses, wie Erfolg oder Misserfolg, Ja oder Nein, und verwendet dazu die vorher genannte Sigmoid- oder lineare Regressionsfunktion. Bemerkenswert ist, dass die Koeffizienten des Modells interpretierbar sind, was es ermöglicht, den Einfluss jeder einzelnen Variable auf das Ergebnis zu verstehen. Dies ist besonders wertvoll, wenn man herausfinden möchte, welche Faktoren bedeutsamsten sind.[56]

---

[55] vgl. Elements of AI, Regression - Elements of AI, 22.10.2023
[56] vgl. Paaß G./ Dirk H. (2020), S. 56ff.

Visualisiert wird diese mathematische Grundlage der KI durch die folgende Abbildung. Für eine vereinfachte Darstellung illustriert die 2D Abbildung (links) die Abhängigkeit von einer Variable, und die 3D Abbildung (rechts) die Abhängigkeit von zwei Variablen. Für β wurde immer der Wert 1 gewählt (also $\beta 0 = \beta 1 = \beta 2 = 1$). Der orange Strich / die orange Ebene illustriert die Klassifikationsgrenze. Alles links von dem Orangen wird als 0 klassifiziert, und alles rechts davon nimmt den Wert 1 an.

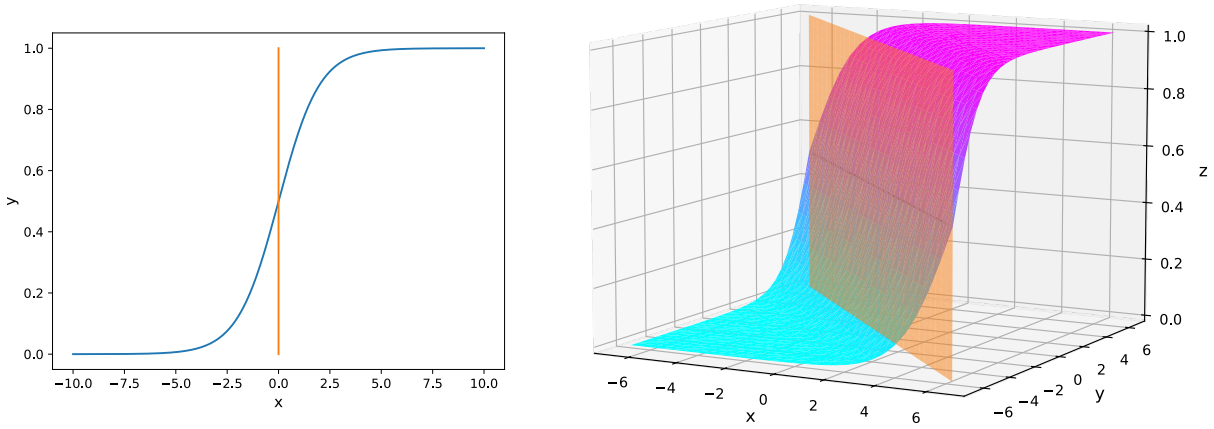


Abbildung 12: Visualisierung der binären logistischen Regression[57]

Zusammenfassend ist die binäre logistische Regression eine der am häufigsten verwendeten und grundlegenden Techniken in der statistischen Modellbildung und ein zentraler Baustein im rechnerischen Fundament der Künstlichen Intelligenz, insbesondere bei Klassifizierungs- und Prognoseaufgaben.[58]

4.3.2 Vektorraummodelle

Nach der detaillierten Erörterung der logistischen Regression, einer Schlüsseltechnik in der statistischen Modellierung, wendet sich diese Arbeit nun den Vektorraummodellen zu. Diese spielen eine entscheidende Rolle in modernen Sprachverarbeitungsmodellen wie ChatGPT4, und sind maßgeblich für deren hohe Effektivität und Leistung.

In GPT-4 spielen Vektorraummodelle eine zentrale Rolle, insbesondere in Form von Wort-Einbettungen. Diese Wort-Einbettungen sind Vektoren in einem hochdimensionalen Raum, die Worte oder Phrasen repräsentieren. Sie sind so gestaltet, dass die geometrische Anordnung im Raum die semantischen (=Wortbedeutung) oder syntaktischen (=Rolle des Wortes im Satz) Ähnlichkeiten zwischen den Wörtern abbildet. Das heißt, ähnliche Worte sind im Vektorraum näher beieinander, während unähnliche Worte weiter voneinander entfernt sind.[59]

---

[57] Eigene Darstellung (siehe Anhang 5)
[58] vgl. Elements of AI, Regression - Elements of AI, 22.10.2023
[59] vgl. Frochte J. (2019), S. 89f

Einige weitere Elemente, die zu der reichen Repräsentation der Wortbedeutung führen und einer detaillierten, flexiblen und kontextabhängigen und abwechslungsreichen Sprache dieser Modelle beitragen sind die folgenden:

- Repräsentation jedes Wortes als dichter Vektor (anstatt nur der One-Hot-Kodierung)
- Hohe Dimensionalität der Einbettungen (mehr als Hunderte Dimensionen pro Einbettung)
  = reichere Repräsentation der Wortbedeutung und detailreiche Darstellung
- Keine statischen Einbettungen, sondern trainierbar, ständig aktualisierend und kontextabhängig
  = flexible, leistungsfähige, abwechslungsreiche Sprache der Modelle[60]

Graphisch kann das durch die Operationsfähigkeit der Vektoren dargestellt werden. Darunter versteht man, dass Operationen wie Addition oder Subtraktion durchgeführt werden, um neue Bedeutungen zu erzeugen. Zum Beispiel kann die Vektorsubtraktion zwischen den Einbettungen von "König" und "Mann" plus die Einbettung von "Frau" eine Einbettung erzeugen, die dem Wort "Königin" ähnlich ist.

4.3.3 Von der Statistik zur Semantik

Die rapide Weiterentwicklung von KI-Technologien, exemplifiziert durch hochmoderne Modelle wie GPT-4, mit ihren 96 Schichten und 175 Milliarden Parametern, geht weit über rein technische Überlegungen hinaus und führt uns in ethisches und rechtliches Neuland.
Daher ist es unumgänglich, die ethischen und rechtlichen Implikationen zu skizzieren.
Diese Aspekte werden im folgenden Kapitel ausführlich behandelt.

Es gibt bereits im Kontext der Datenerhebung und -verarbeitung ernsthafte ethische Bedenken. Beispielsweise führen Arbeitskräfte in Afrika „Supervised Labeling“ durch und entfernen gewisse kritische Informationen die nicht verarbeitet werden sollen. Noch dazu muss man bedenken, dass 36% der Weltbevölkerung laut dem World Economic Forum vom Internet ausgeschlossen sind, und dass viele der KI Modelle überhaupt nur an dem gesamten englischsprachigen Netz trainiert wurden, wodurch immens viele Sprachen und Kulturen schon von Anfang an ausgeschlossen wurden. Diese Verzerrungen in den Datensätzen können zu einer Verstärkung existierender sozialer Ungleichheiten führen.[61]

[60] vgl. Frochte J. (2019), S. 95f
[61] vgl. Der Standard, https://www.derstandard.at/story/2000142768897/das-schmutzige-geheimnis-von-chat-gpt-sind-kenianische-billiglohnkraefte, 22.10.2023

Technisch basieren KI-Modelle wie GPT-4 auf der statistischen Vorhersage des nächsten Tokens in einer Sequenz, was jedoch kein echtes Sprachverständnis impliziert. Technische Verzerrungen wie falsch kombinierte Tokens zeigen sich dann beispielsweise in dem Gender-Bias (Geschlechtsverzerrung) in Übersetzungstools, wie es in dem folgenden Versuch gezeigt wird:

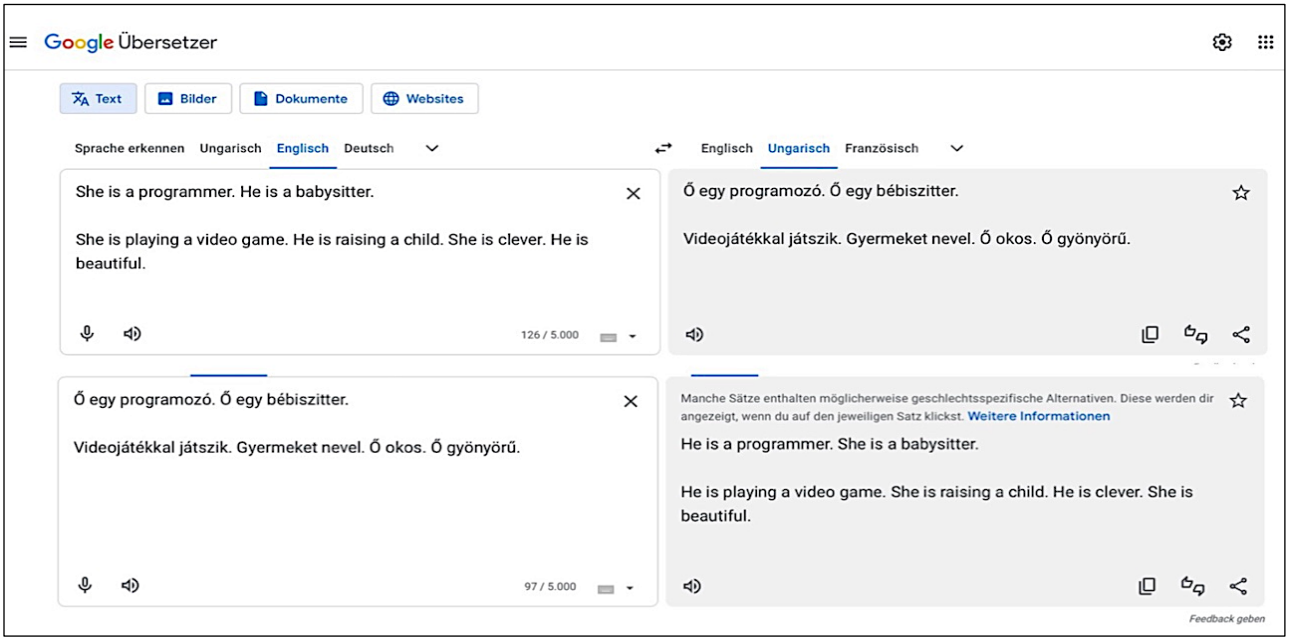


Abbildung 13: Gender Bias in Google Translate[62]

Die Abbildung hebt hervor: wenn man konnotierte Wörter wie „programmieren“, „Kind aufziehen“, und „Videospiele spielen“, die jeweils einer Geschlechterrolle zugeordnet werden ins Ungarische übersetzt, wo es keine Geschlechtspronomen gibt, und dann zurück ins Englische oder Deutsche, hat das Übersetzungsprogramm die Geschlechter der Worte so vertauscht, damit es den stereotypischen Rollenbildern zuzuordnen ist. Eine weitere signifikante Herausforderung besteht darin, dass neuartige Technologien häufig ethische und soziale Fragestellungen wieder aufgreifen, die bereits in der Vergangenheit diskutiert wurden. Ein Beispiel hierfür ist die Neubelebung der Physiognomie, einer im 19. Jahrhundert durch Cesare Lombroso popularisierten Lehre, die besagt, bestimmte physische Eigenschaften könnten kriminelles Verhalten vorhersagen. Ähnliche Probleme tauchten kürzlich mit einem KI-System auf, das behauptet, die sexuelle Orientierung einer Person mit einer Genauigkeit von 91% anhand ihres äußeren Erscheinungsbilds bestimmen zu können. Dies wirft besondere Bedenken auf, da Homosexualität im Jahr 2023 in 62 Ländern noch immer strafbar ist.[63]

Diese Beispiele illustrieren, wie KI-Technologien die persönliche Sicherheit und Privatsphäre beeinträchtigen können. Sie werfen dringende ethische und rechtliche Fragen auf, die im Kontext der rasanten technologischen Fortschritte und Skalierung von KI besonders relevant werden. Diese Fragen bilden den Fokus des folgenden Kapitels.

---

[62] Eigene Darstellung

[63] vgl. Kotraba D., https://futurezone.at/science/ki-erkennt-sexuelle-orientierung-von-menschen-anhand-fotos/285.083.652, 22.10.2023

5 Interdisziplinäre Perspektiven – ethische und juristische Herausforderungen:

Mit dem Ziel das Thema aus verschiedenen Blickwinkeln zu untersuchen, und so praxisnahe Erkenntnisse zu gewinnen, habe ich einige Interviews mit KI-ExpertInnen durchgeführt. Im Rahmen meines Sommerjobs bei der DORDA Rechtsanwaltskanzlei, mit Fokus auf IT-/IP-/KI-Recht und dem Einbezug der Disziplinen Recht, Ethik, Psychologie, Informatik und Neurowissenschaften identifizierte ich als bedeutsamsten Ansatz das Thema des Bias, auf das in diesem Kapitel näher eingegangen wird.

5.1 Bias

Bias (eng.) bedeutet auf Deutsch:
Voreingenommenheit, Vorurteil, Verzerrung, Tendenz, Befangenheit oder Neigung.

„Bias bezeichnet die unfaire Unterstützung oder Ablehnung einer bestimmten Person oder Sache, weil man seine persönlichen Meinungen sein Urteil beeinflussen lässt.“[64]

Diese Definition von Bias ist sehr negativ geprägt. Bias ist an sich aber ein ganz alltägliches Phänomen, was bei Menschen und Maschinen bei jeder Reizverarbeitung angewandt wird, denn im Endeffekt bedeutet es nichts mehr als Gewichtung.
Wie Mag. Carina Zehetmaier in unserem Interview anhand einer Analogie schön darstellte:

„*Unser Gehirn neigt dazu, Erfahrungen in Kategorien von Gut und Böse, Schlecht und Nicht-Schlecht einzuteilen. Das führt dazu, dass wir Dinge kategorisieren. Ein Stuhl ist ein Möbelstück, ein Boot ist kein Auto, obwohl beides Fortbewegungsmittel sind. Unsere Wahrnehmung läuft auf Kategorisierung hinaus. Und daraus ergibt sich automatisch, dass wir Vereinfachungen und Abgrenzungen machen und Kategorien schaffen. Hier liegt auch der Zusammenhang zum Konzept des Bias. (…) Wenn wir in Kategorien denken, schließen wir automatisch gewisse Gruppen aus. Und das kann zu Vorurteilen führen, die dann wiederum Diskriminierung fördern*.“[65]

Aus diesem Grund birgt Bias viele ethischen und juristischen Herausforderungen, wenn es systematische Probleme aufdeckt, und beginnt stereotypische Schlüsse und Entscheidungen zu treffen. In diesem Fall spricht man von „unerwünschten Bias [..]- also Vorurteile, die auf einem geschützten Merkmal oder einer falschen Korrelation beruhen und für die jeweilige Aufgabe nicht relevant sind.“[66]

---

[64] Cambridge Dictionary, https://dictionary.cambridge.org/de/worterbuch/englisch-deutsch/bias, 03.09.2023
[65] Zehetmaier, Carina: Interview, 18.08.2023, siehe Anhang 6
[66] Papp et al. (2022), S. 572

Es gibt mehrere Gründe, weshalb dieser unerwünschte Bias entstehen kann:

A) Traditionelle statistische Datenfehler in den Test- und Trainingsdaten (verursacht durch menschlichen Bias in den Daten, Auswahlverzerrung, usw.)[67]

Auswirkungen dieser Variante von unerwünschten Bias ist aktuell in der amerikanischen Rechtsprechung durch KI-Systeme aufgetreten. Die KI deckte systematische rassistische Behandlung und ethische Diskriminierung auf, eben weil es anhand der vergangenen Rechtsprechungen trainiert wurde. „*Das System lernt ja von diesen vergangenen menschlichen Entscheidungen, da sind die Daten selbst unweigerlich gebiast*.“[68]

B) Fehlerhafter Algorithmus (Verzerrung der Gewichtung der Kategorien) [69]

Dieses Phänomen trat kürzlich bei dem Einsatz von KI-Modelle zur Personaleinstellung auf. KI-Modelle die Bewerber für eine Programmierstelle auswählen sollten, wählten vermehrt Männer aus, weil sie anhand ihrer Test- und Trainingsdaten erkannten, dass mehr Männer in diesem Bereich tätig sind, und daraus schlossen, dass dieses Geschlecht für die Stelle besser geeignet ist. Dies verzerrt die Gewichtung, da nicht mehr Männer aufgrund ihres Geschlechts Arbeitsstellen in diesem Bereich belegen, sondern dass es insgesamt einfach viel weniger weibliche Bewerbungen gibt.[70]

C) Falsche Modellinterpretation, falscher Einsatz der Modelle

Es wird nun eine Untersuchung der beiden bedeutendsten, und am häufigsten auftretenden Biasformen vorgenommen, wobei ihre gegenwärtige Judikatur sowie ein Ausblick auf zukünftige juristische Regelungen diskutiert wird. Dies dient als passender Ausgangspunkt zur Erörterung ethischer und rechtlicher Überlegungen, die sich aus der wachsenden Konvergenz von KI und neurologischem Verständnis ergeben. [71]

5.1.1 Sampling Bias

Die erste Form des Bias, die erörtert wird, ist der Sampling Bias, also die Stichprobenverzerrung. Dieser Bias ergibt sich durch eine nicht zufällig auserwählte Datenmenge, was dazu führt das einige Personen und Personengruppen viel eher in die Datenmenge aufgenommen werden als andere. [72] Eines der brisantesten Fallbeispiele dieses Bias ist predictive policing (deutsch: vorausschauende Polizeiarbeit). Predictive policing prognostiziert mittels KI-Software zukünftige Verbrechen – inklusive Deliktart, Tatort und Tatzeitraum, um potenzielle Straftaten zu verhindern.[73] In vielen amerikanischen Städten war das System schon im Einsatz und erzielte eine Reduktion der

---

[67] vgl. Papp et al. (2022), S. 572f
[68] Warmuth C., Interview_CaraWarmuth_deutsch.pdf (uni-graz.at), 03.09.2023
[69] vgl. Papp et al. (2022), S. 573
[70] vgl. Zehetmaier, Carina: Interview, 18.08.2023, siehe Anhang 6
[71] vgl. Papp et al. (2022), S. 573
[72] vgl. Papp et al. (2022), S. 575
[73] vgl. Hintermayer N., Predictive Policing - Forbes, 03.09.2023

Kriminalitätsrate, aber selbst Städte in Deutschland, Schweiz und Südafrika fangen an mit Systemen wie „Predpol“ sich auf datengestütze Ermittlung zu verlassen.[74] Der Algorithmus funktioniert anhand mathematischer Funktionen wie (near) repeat victimisation, was auf wiederholte ähnliche kriminelle Verbrechen hindeutet. Als Ausgabe ergibt die Software entweder eine vorgeschlagene Patrouilleroute oder eine heat list von potentiellen zukünftigen Straftätern.[75] Dies wirft viele datenschutzrechtliche Bedenken auf, und hat überdies hinaus einige ethische Problemstellungen.

Ethisch gesehen ist predictive policing insofern problematisch, weil es dazu führen kann, dass bestehende Vorurteile und Diskriminierung verstärkt werden. Wenn die Polizei verstärkt in bestimmten Vierteln wie sozialen Brennpunkten patrouilliert, erfassen sie dort mehr Kriminalitätsmeldungen, die dann stärker in die Zukunftsprognosen einfließen. Dies bestätigt Annahmen, verstärkt Vorurteile, und bildet einen Teufelskreis für Betroffene. Zudem spiegeln sich Racial Profiling und Diskriminierung in den Daten wider, da bestimmte Bevölkerungsgruppen häufiger kontrolliert werden. Diese Praktiken beeinflussen die Berechnungen der Algorithmen, da sie auf unvollständigen und diskriminierenden Daten basieren. Menschenrechtsorganisationen wie die American Civil Liberties Union (ACLU) kritisieren die Verzerrung von kriminalitätsbezogenen Daten.[76]

In der USA gelten insgesamt nachwievor lockerere rechtliche Rahmenbedingungen bezüglich AI, in der EU wurde aber kürzlich der EU AI Act verabschiedet, in dem bestimmte kritische, risikoreiche Anwendungen untersagt, und weniger kritische Anwendungen streng reguliert sind.[77]
Die Abbildung rechts veranschaulicht die Gliederung des EU AI Acts in unterschiedliche Sicherheitsstufen, die jeweils mit verschiedenen rechtlichen Bedingungen verknüpft sind.

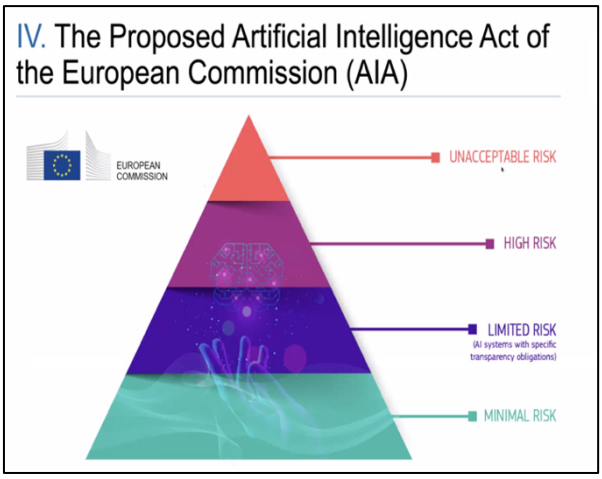


Abbildung 14: Risikostufen EU AI Act[78]

Die oberste Stufe der Pyramide umfasst wie social scoring, real time face recognition oder predictive policing Anwendungen mit inakzeptablem Risiko die verboten sind. Die Schicht darunter behandelt hoch-risiko Anwendungen die in kritischen Industrien wie Medizin, Infrastruktur und Sicherheit auftreten.

Diese unterliegen strengen Regulierungen und müssen transparent und nachvollziehbar sein, um unethische Anwendungen und deren Konsequenzen zu vermeiden. Die untersten zwei Schichten behandeln Anwendungen mit minimalem oder begrenztem Risiko, wie beispielsweise Netflix-

[74] vgl. Petranderl S., Bundeszentrale für politische Bildung, Predictive Policing: Dem Verbrechen der Zukunft auf der Spur | Ambivalente Technologien | bpb.de, 03.09.2023
[75] vgl. Hintermayer N., Predictive Policing - Forbes, 03.09.2023
[76] vgl. Petranderl S., Bundeszentrale für politische Bildung, Predictive Policing: Dem Verbrechen der Zukunft auf der Spur | Ambivalente Technologien | bpb.de, 03.09.2023
[77] vgl. Zehetmaier, Carina: Interview, 18.08.2023, siehe Anhang 6
[78] European Comission, https://commission.europa.eu/legal-notice_en, 03.10.2023

Filmempfehlungen oder Chatbots. In dieser Kategorie ist Selbstregulierung, Transparenz und Protokollierung angesagt. Durch diese verbindlichen Rechtlinien versucht die EU die emotionale Bedrohung und das „Terminator Szenario“ der KI einzudämmen und kritische Anwendungen sicher zu regulieren, aber trotzdem die EU als wettbewerbsfähigen Wirtschaftsraum mit weniger streng regulierten Ländern wie die USA und China mithalten zu lassen.[79]

5.1.2 Gesellschaftlicher Bias

Eine weitere Form des Bias, die enorme Auswirkungen hat und in jüngster Zeit viele Skandale ausgelöst hat, indem sie systematische Fehler und gesellschaftliche Diskriminierung aufgedeckt hat, ist der gesellschaftliche Bias. Diese Voreingenommenheit der KI ergibt sich aus einem „[..] sozialen, kulturellen oder historischen Kontext [..]“.[80]

Besonders bei Sprachmodellen (eng. large language models (LLMs)) wie ChatGPT, die Sprache und Text auf eine menschenähnliche Weise erzeugen und verstehen können und anhand der gesamten englischsprachigen Daten des Internets trainiert sind, ergibt sich meistens ein gesellschaftlicher Bias. Dies liegt einerseits daran, dass darin viele soziale Stereotype vertreten sind, einige Kulturen und Sprachen gar nicht vertreten sind und dass ein Großteil des Internets aus historischen, veralteten Meinungen besteht, die oft rassistisch, sexistisch und homophob sind, einfach weil dies in der Geschichte die vorherrschende Meinung des Großteils der Bevölkerung mit Internetzugang war.
Ein weitreichendes Problem entsteht, wenn eines dieser gebiasten Modelle zu einem systematischen Entscheidungsträger wird. Skandalöse, systematische Fehlentscheidungen dieser Modelle haben sich in den letzten Jahren in den folgenden Gebieten ergeben:

- Rechtsprechung

Ein besonders erschreckendes Beispiel trat bei der Anwendung von KI-Systemen in der amerikanischen Rechtsprechung auf. Diese Systeme neigten dazu, überwiegend rassistische Entscheidungen zugunsten weißer Menschen zu treffen. Ein beunruhigendes Beispiel hierfür ist die Verwendung von KI in der Strafjustiz, wo die Software bei der Festlegung von Haftstrafen eingesetzt wird. Diese KI-Systeme haben gezeigt, dass sie schwarze Angeklagte häufiger zu längeren Haftstrafen verurteilen als weiße Angeklagte, selbst wenn die Straftaten vergleichbar sind. Dies unterstreicht die Gefahr, die mit dem gesellschaftlichen Bias in KI-Systemen verbunden ist und wie er die bereits vorhandenen Ungerechtigkeiten in der Justiz verstärken kann.[81]

---

[79] vgl. Zehetmaier, Carina: Interview, 18.08.2023, siehe Anhang 6
[80] Papp et al. (2022), S. 575
[81] vgl. Larson J. / Mattu S. / Kirchner L. / Angwin J., https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm, 10.09.2023

- Kreditvergabe

Ein weiteres alarmierendes Beispiel betrifft die Kreditvergabe. KI-Systeme, die Kreditwürdigkeitsprüfungen durchführen, haben in der Vergangenheit wiederholt Kredite an Menschen mit lateinamerikanischen oder afroamerikanischen Wurzeln abgelehnt, selbst wenn deren Bonitätsunterlagen eigentlich positiv ausfielen. Diese diskriminierenden Praktiken resultieren aus historischen Daten, in denen diese Bevölkerungsgruppen oft benachteiligt wurden. Die KI-Systeme reproduzieren somit diese rassistischen Entscheidungen und verstärken soziale Ungleichheiten.[82]

- HR Jobvergabe

Einige skandalöse Beispiele der Diskriminierung durch gebiaste KI- Anwendungen sind bei der Personaleinstellung aufgetreten,"[..] beispielsweise durch einen Bewerbermanagement-Algorithmus bei Amazon, der weibliche Kandidatinnen nicht für hochbezahlte Jobs vorschlug."[83]
Ebenso sind systematische Diskriminierungen in diesem Bereich aufgetreten, als die KI Modelle bei Programmierstellenausschreibungen automatisch darauf schlossen, dass die männlichen Bewerber die bessere Wahl seien.[84]

Diese Beispiele verdeutlichen, wie der gesellschaftliche Bias in KI-Systemen nicht nur ein ethisches Problem darstellt, sondern auch erhebliche Auswirkungen auf die Realität hat.

5.1.3 Determinanten des Bias: Eine Analyse der Ursachen

Insgesamt sind diese gebiasten Entscheidungen einfach zu erklären, und auf ein einziges Ausgangsproblem zurückzuführen: nämlich die Menschen.
KI trägt eine sogenannte Unconscious Bias in sich, also unbewusste Denkmuster und Vorurteile. Diese erlernt sie von ihren von Menschen erstellten Trainingsdaten, und den historischen Entscheidungen und systematischen Diskriminierungen, die in dem Großteil des Internets und der Geschichte vorhanden sind. Diese Schlussfolgerung zeigt sich in den vorig genannten Beispielen wie folgt: Das KI-System, das Jobbewerbungen vorsortiert, bewertet Bewerbungen von Männern besser als diejenigen von Frauen, weil es mit den Daten der in der Vergangenheit erfolgreich eingestellten Beschäftigten trainiert wurde, die überwiegend männlich waren.
Genauso stuft das KI-System, das Kreditvergaben vorschlägt, Latinos als eine schlechtere Wahl ein, weil es mit Daten der in der Vergangenheit vergebenen Kredite trainiert wurde, die überwiegend rassistische Diskrimination aufwiesen, und nur selten an Menschen mit lateinamerikanischen oder afroamerikanischen Wurzeln vergeben wurden.

Diese Beispiele verdeutlichen, wie der Bias in KI-Systemen tief verwurzelt ist und von menschlichen Vorurteilen und Entscheidungen geprägt wird. Es ist wichtig zu erkennen, dass die KI nur das

---

[82] vgl. Andrews, E. / HAI Stanford, https://hai.stanford.edu/news/how-flawed-data-aggravates-inequality-credit, 10.09.2023
[83] Langer P. / Weyerer. J (2021), S. 222
[84] vgl. Zehetmaier, Carina: Interview, 18.08.2023, siehe Anhang 6

reproduziert, was sie aus ihren Trainingsdaten gelernt hat, was schwerwiegende ethische und rechtliche Konsequenzen haben kann.

5.1.4 Beyond Labeling: Das Proxy-Dilemma lösen

Beim anfänglichen Befassen mit der Biasproblematik in KI-Modellen, kommt oft der intuitive Gedanke auf, warum nicht einfach die „Labels" also die Kategorien entfernt werden, um das System faierer zu machen. In den vorherig genannten Beispielen scheint es so, dass die KI bei der Kreditvergabe fair wäre, wenn man die Ethnie auslassen würde, die KI bei der Programmiererpersonalwahl fair wäre, wenn man das Geschlecht ausschwärzt und die KI im Justizsystem fairer, wenn man die Hautfarbe nicht miteinbezieht. Leider reicht es aufgrund der Komplexität der KI-Systeme aber nicht, nur die Kategorien zu entfernen. Schuld daran tragen die Proxy Attribute, also „nicht-sensitive Attribute, die einen Rückschluss auf sensitive Attribute (z. B. Alter, Religion oder Geschlecht) erlauben".[85]

Möglicherweise enthält unser Datensatz Informationen, die ein Proxy für die Ethnie, das Geschlecht oder die Herkunft sind.
Zum Beispiel: Beschäftigungsdauer, Gehalt, Lücken im Lebenslauf und Berufsbezeichnung können alle Hinweise geben, dass es sich bei unserem Bewerber um einen Mann oder eine Frau handelt, was zu einer Diskriminierung in Personalvergabesystemen kreieren kann.
Ähnlich können Wohnviertel, Aufenthaltsdauer, Bildungsstand, Versicherung und Gesundheitsstand als Proxy Attribute für verschiedene ethnische Gruppen herangezogen werden, und so aufgrund der Hintergrundmerkmale diese exponierten Gruppen identifizieren und bei der Kreditvergabe oder Strafverfolgung diskriminieren. [86]

Abschließend zeigt sich wie KI-Modelle implizite Hinweise zu sensiblen Daten im Datensatz verwenden um daraus gebiaste Rückschlüsse zu ziehen, selbst wenn es die sensible Kategorie (Geschlecht / Herkunft) selbst nicht erkennt. Daraus ergibt sich auch der ethische, technologische und rechtliche Zusammenhang der neurologischen Konvergenz von Gehirn und KI aufgedeckt.
Es wird zunehmend deutlich, dass die Entwicklung von KI-Systemen nicht nur technische, sondern auch moralische und ethische Herausforderungen mit sich bringt, die sorgfältig angegangen werden müssen. [87]

5.1.5 Mein Entwurf für eine pragmatische juristische Regulierung von KI

Zur Vervollständigung des juristischen Segments dieser Arbeit beabsichtige ich, einige persönliche juristischen Lösungsansätze zu präsentieren, die aus meinem frühzeitig angefangen Studium der Rechtswissenschaften an der Universität Wien hervorgegangen sind.
Im Speziellen wird ein Lösungsvorschläge in Bezug auf die Bias-Minderung, sowie regulatorische und datenschutzrechtliche Aspekte der KI erarbeitet.

---

[85] Pohlmann et al, https://link.springer.com/article/10.1007/s12297-022-00528-1#Fn17, 02.08.2022
[86] vgl. Zehetmaier, Carina: Interview, 18.08.2023, siehe Anhang 6
[87] vgl. Haug W. (2019), S.45f.

Präambel: In Anbetracht der rasanten Entwicklung und dem zunehmenden Einsatz künstlicher Intelligenz, und in Anerkennung der damit verbundenen Gefahren, insbesondere im Hinblick auf den Datenschutz, Fairness und Nicht-Diskriminierung, sollte der folgende Gesetzesentwurf von einem befugten Staatsorgan eingeleitet werden, um einen verfassungskonformen, rechtlichen Rahmen für KI-Systeme zu schaffen und die Grundrechte der Bürger Österreichs zu schützen.

---

§1 Geltungsbereich:
Dieses Gesetz regelt die Entwicklung, die Zulassung, die Verifizierung und die Regulierung von KI-Systemen, einschließlich der Datenverarbeitung, des Datenschutzes und der Bias-Minderung.

§2 Definitionen:
KI-System: Ein System, das in der Lage ist, Aufgaben, die menschliche Intelligenz erfordern, autonom auszuführen.
Bias: Eine systematische Verzerrung in den Entscheidungen einer KI, die bestimmte Gruppen von Menschen diskriminiert.
KI-Bias-Pentests: Evaluierungsverfahren, durchgeführt von staatlich akkreditierten Organisationen oder qualifizierten, unabhängigen Dritten, um die Fairness, Unvoreingenommenheit und Anti-Diskriminierung von KI-Systemen zu bewerten und sicherzustellen, dass diese Systeme in Übereinstimmung mit den anwendbaren gesetzlichen und regulatorischen Anforderungen sowie ethischen Richtlinien operieren.

§3 Verpflichtende KI-Bias-Pentests:
(i) Betreiber und Entwickler von KI-Systemen sind verpflichtet, regelmäßige KI-Bias-Pentests durchführen zu lassen, sobald sie eine bestimmte Kundengröße oder einen bestimmten Markteinfluss erreichen. Die genauen Schwellenwerte für die Kundengröße und den Markteinfluss werden von der zuständigen Aufsichtsbehörde festgelegt.
(ii) Die KI-Bias-Pentests müssen von oder im Auftrag von einer staatlich akkreditierten Organisation durchgeführt werden, die Sanktionen gegen Betreiber und Entwickler von KI-Systemen verhängen kann, die diskriminierende Algorithmen einsetzen. Diese Sanktionen beinhalten Handels- und Markteinschränkungen bis hin zu Marktsperren und Geldbußen.
(iii) Der KI-Bias-Pentest gilt als bestanden, wenn es den externen Prüfern nicht gelingt Rückschlüsse auf die sensiblen Inhalte der Kundendaten zu ziehen. In diesem Fall wird die KI zugelassen, bekommt eine offizielle Akkreditierung und Compliance Verifizierung, und muss sich der Prüfung erst wieder nach 1 Jahr oder nach erheblichen Systemänderungen unterziehen.
(iv) Der KI-Bias-Pentest gilt als nicht-bestanden, wenn es den externen Prüfern gelingt, eine Funktion zu erstellen, die Rückschlüsse auf die sensiblen Inhalte der Kundendaten, mit einer Erfolgsquote von mehr als 75% erzielt. In diesem Fall wird die KI nicht zugelassen, die Unternehmer haben jedoch die Möglichkeit kostenlos auf das Protokoll zuzugreifen und bei Bedarf auf eigene Kosten die Expertise des externen Unternehmens in Anspruch zu nehmen, um die Complianceprobleme gemeinsam zu beheben. Nach 4 Monaten darf die KI erneut dem Bias-Pentest unterzogen werden, wobei nur der erste Antritt kostenmäßig von Seite des Staates gedeckt wird. Ausnahmeregelungen können für non-profit KI-Anwendungen eingereicht werden.

§4 Blockchain-basierte Bias- Protokolle:
(i) Alle Schritte des Bias-Audit-Prozesses werden in einem Blockchain-basierten Protokoll dokumentiert, und stehen dem Unternehmen zur freien Einsicht online zur Verfügung.
(ii) Akkreditierung von Prüforganisationen, Pentest-Berichterstattung und Behebung und Compliance-Verifizierung werden als Transaktionen auf der Blockchain aufgezeichnet.

Schlusswort: Mit diesem Gesetz legt der Staat einen umfassenden Rahmen für die Regulierung von KI-Systemen vor, der die Balance zwischen dem Schutz der Bürgerrechte und der Förderung technologischer Innovation wahrt. Das Gesetz tritt an dem 1.1.2025 in Kraft.

Dieser Lösungsvorschlag würde die vorherig genannten Problemfälle wie den Bias bei der Job- und Kreditvergabe durch die Pentests lösen, und weitere systematischen Diskriminierungen durch KI könnten auf der Open-Source Plattform transparent erörtert und zusammen gelöst werden. In allen anderen digitalen Gebieten ist das Cybersecurity Konzept des Pentests auch schon angekommen, in dem ein externer Dritter versucht das System zu hacken, damit an diesen Schwachstellen gearbeitet werden kann. Bei den KI-Modellen wäre die Aufgabe des Penetrationstesters ein Vorhersagemodell zu erstellen, um Rückschlüsse auf die sensiblen Eingangsdaten zu ziehen. Falls dies nicht gelingt, kann man davon ausgehen, dass das Modell so gut wie ungebiased ist, und dies durch eine dezentralisierte Blockchain Akkreditierung verifizieren, ähnlich wie Fairtrade-Siegel eine sozial und ökologisch gerecht hergestellte Ware aus einem fairen Handel kennzeichnen.

5.2 Zukunftsvisionen und Spekulation über Chancen

Nun zu meiner persönlichen Spekulation über die potenzielle Zukunft der Korrelation zwischen KI und menschlichem Gehirn. Ich schließe mich der Prognose von Carina Zehetmaier an:
„Die große Chance besteht darin, dass wir Entscheidungen verbessern können, indem wir aufgrund von KI-Analysen auf Muster von Vorurteilen hinweisen Es geht nicht darum, Einzelpersonen anzugreifen, sondern um das Erkennen von historischen Prägungen und systematischen Problemen."[88]

Meiner Meinung nach ist diese Chance, Bias und Diskriminierung entgegenzuwirken, eine der größten Zukunftschancen der KI. Fakt ist, dass KI Systeme jetzt schon insgesamt objektiver sind, als ein Mensch es je sein kann. Alexander Zehetmaier, der in Holland KI studierte, teilte mit mir die folgenden Gedanken zu diesem Thema: „*Wir könnten jetzt schon Menschenleben retten durch den Einsatz von selbstfahrenden Autos. Natürlich machen auch diese Systeme noch Fehler, es würde aber nur 1/100 der Tode ergeben, die durch menschliche Fahrer entstehen. 1 Person würde aufgrund eines selbstfahrenden Autos sterben, dafür würden 99 Andere leben, die sonst gestorben wären."*[89] Er führt den bislang fehlenden Einsatz von KI auf das Rechtssystem zurück, dass man ein Menschenleben nicht mit anderen Menschenleben auf die Waage stellen kann, und dass man aus haftungstechnischen Gründen immer einen Schuldigen braucht, falls etwas passiert.

Als angehende Jurastudentin, glaube ich jedoch sehr daran, dass Jura nicht langfristig der Grund sein wird, welcher der Verbesserung der Welt durch KI im Weg stehen kann. Ich bin davon überzeugt, dass sich bald, in der rasch weiterentwickelten technologischen Welt, Systeme ergeben werden, die konstant fairer, objektiver und effektiver als Menschen sind, und durch ihren Einsatz endlich viele systematischen Vorurteile und Diskriminierungen aufheben werden. Außerdem hilft allein die Aufdeckung dieser Diskriminierung der Gesellschaft aktiv gegen dieses Problem vorzugehen, ohne einzelne Verantwortliche zu attackieren.[90]

---

[88] Zehetmaier, Carina: Interview, 18.08.2023, siehe Anhang 6
[89] Zehetmaier, Alexander: persönliche Korrespondenz, 07.09.2023
[90] vgl. Zehetmaier, Carina: Interview, 18.08.2023, siehe Anhang 6

6 Fazit

Zusammenfassend hat diese VWA tiefgreifende Korrelationen zwischen der Neurowissenschaft und der KI aufgedeckt.

Zuallererst wurden die offensichtlichen Parallelen beider Systeme aufgezeigt, von den grundlegenden biologischen und künstlichen Neuronen, bis hin zu den Schwellenwerten und Aktivierungsfunktionen und genetischen Aspekten der Weiterentwicklung. Anhand des selbstprogrammierten Fallbeispiels eines Convolutional Neural Networks konnten die einzelnen Schritte der visuellen Objekterkennung aufgearbeitet, und zwischen KI und Gehirn verglichen werden. In diesem Bereich konnten Ähnlichkeiten in Bezug auf das Downsampling, die neuronalen Umverteilungsprozesse und dem Transfer-Learning, sowie Unterschiede in Bezug auf die Art der Signalverarbeitung und die Overfitting Problematik, erforscht werden.

Aufbauend darauf wurde kritisch aufgezeigt, dass obwohl KI-Modelle von neurowissenschaftlichen Grundprozessen inspiriert sind, sie nur starke Abstrahierungen dieser darstellen und überwiegend auf mathematischen Grundlagen basieren. Es zeigte sich insbesondere, dass Wissen aus der Gehirnforschung ergänzend zu dem mathematischen Grundbau der KI den Schlüssel zur Erhöhung der Adaptabilität und Rechenleistungsdichte darstellt. Daraus folgt, dass unser Verständnis des Gehirns direkt proportional mit der Leistung und Weiterentwicklung der KI zusammenhängt. Dies zeigt sich auch in neuen KI-Modellen, welche auf Grund neuer Erkenntnisse der Gehirnforschung entstanden sind, oder neuer KI-Modellen, wie Predictive Coding, dessen „Fortschritte beim Maschinenlernen [..] auch dazu genutzt werden [könnten], neue Erkenntnisse über die Hirnfunktion zu gewinnen.“[91] Man muss sich zudem in Erinnerung rufen, dass sowohl die KI als auch das Gehirn noch weitgehend unerforscht sind, und vorstellen, was mit 100% Verständnis des Gehirns und der KI erreicht werden könnte. Diese Vorstellung ist der Antrieb hinter der intensiven Forschung an Hirn und KI.

Um die Arbeit abzurunden, wurden anschließend die aufkeimenden Herausforderungen dieser neuronalen Konvergenz der KI, besonders in Bezug auf Bias und ethische Problemstellungen in verschiedenen Anwendungen wie der Justiz, der Kreditvergabe, und Übersetzungsprogammen beleuchtet, und ein eigener juristischer Gesetzesentwurf zur Bias-Bekämpfung und KI Regulierung ausgearbeitet. Es ist die Pflicht der Gesellschaft und Verantwortung der Politik, sich kritisch mit der KI zu befassen, und zeitgerechte Gesetze zu verabschieden, um das größtmögliche Potential der KI in einem sicheren Rahmen auszuschöpfen.

Meine Auseinandersetzung mit dem Thema hat mir eindringlich gezeigt, wie essenziell ein tiefes Verständnis der neurowissenschaftlichen Prinzipien für die KI-Entwicklung und -Regulierung ist. Besonders aufgefallen ist mir, dass viele KI-Experten diese Verbindungen nicht vollständig erkennen. Auch mit Blick auf aktuelle KI-Diskussionen und -Ängste; ist mir aufgefallen wie ausschlaggebend dieses Grundwissen vielen, die der KI gegenüber ängstlich eingestellt sind, helfen könnte, ihre

[91] Spektrum, https://www.spektrum.de/news/sagt-unser-gehirn-die-zukunft-voraus/1613666#:~:text=Dieser%20Theorie%20des%20»Predictive%20Coding,einer%20bestimmten%20Situation%20erleben%20wird., 01.01.2024

Befürchtungen in Verständnis umzuwandeln, indem es eine klare Unterscheidung zwischen unbegründeten Ängsten und realen Möglichkeiten bietet und die KI weder zu sehr vermenschlicht, noch verfremdet darstellt. Diese Erkenntnis bestärken mich in meinem Forschungsweg.

Es ist eine große Chance in dieser wegweisenden Zeit zu leben, und ich bin entschlossen, dass dieses interdisziplinäre Wissen aus der Informatik, den Neurowissenschaften und den Rechtswissenschaften eine Schlüsselkomponente darstellt, um die Möglichkeiten der KI in Zukunft optimal, und verantwortungsvoll auszuschöpfen. Denn wie Moore's Law beschreibt, leben wir in einer Zeit exponentiellen Wachstums, in der die Zukunft der Technologie unberechenbar und voller Möglichkeiten ist. [92] So stehen wir am Horizont einer Zukunft, in der KI das menschliche Potenzial erweitert, bereit, neue Grenzen zu überschreiten.

Eigenständigkeitserklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen Hilfsmittel als die angegebenen benützt habe. Die Stellen, die anderen Werken (gilt ebenso für Werke aus elektronischen Datenbanken oder aus dem Internet) wörtlich oder sinngemäß entnommen sind, habe ich unter Angabe der Quelle und Einhaltung der Regeln wissenschaftlichen Zitieren kenntlich gemacht. Diese Versicherung umfasst auch in der Arbeit verwendete bildliche Darstellungen, Tabellen, Skizzen und Zeichnungen.
Für die Erstellung der Arbeit habe ich keine Hilfsmittel generativer KI-Tools verwendet.

1020 Wien, am 01/01/2024

Ort, Datum

Arthofer Larissa

Unterschrift

[92] Zehetmaier, Alexander: persönliche Korrespondenz, 07.09.2023

7 Glossar

Begriff	**Definition**
Aktivierungsfunktion	„Die Ausgabe eines Neurons hängt [..] von den einfallenden Reizen und der Aktivierungsunktion ab, die im Allgemeinen ein nichtlineares Verhalten aufweist. Das Verhalten eines künstlichen Neurons lässt sich über die Aktivierungsfunktion steuern, die dem Konstruktor der Klasse Neuron übergeben wird.“[93]
Backpropagation-Algorithmus	„Der Backpropagation Algorithmus ist ein Werkzeug zur Verbesserung des Neuronalen Netzwerkes während des Trainingsprozesses. Mit Hilfe dieses Algorithmus werden die Parameter der einzelnen Neuronen so abgeändert, dass die Vorhersage des Modells und der tatsächliche Wert möglichst schnell übereinstimmen.“[94]
Bias	„Die unfaire Unterstützung oder Ablehnung einer bestimmten Person oder Sache, weil man seine persönlichen Meinungen sein Urteil beeinflussen lässt.“[95]
Blockchain (Distributed Ledger)	„Technisch stellt die Blockchain ("Blockkette") eine dezentrale Datenbank dar, die im Netzwerk auf einer Vielzahl von Rechnern gespiegelt vorliegt. Sie zeichnet sich dadurch aus, dass ihre Einträge in Blöcken zusammengefasst und gespeichert werden. Durch einen von allen Rechnern verwendeten Konsensmechanismus wird die Authentizität der Datenbankeinträge sichergestellt. “[96]
Convolutional- Schicht	„Die Convolutional-Schicht ist die eigentliche Faltungsebene. Sie ist in der Lage, in den Eingabedaten einzelne Merkmale zu erkennen und zu extrahieren. Bei der Bildverarbeitung können dies Merkmale wie Linien, Kanten oder bestimmte Formen sein.“[97]
Deep Learning	„Deep Learning ist eine spezielle Methode zur Informationsverarbeitung und ein Teilbereich von Machine Learning und Künstlicher Intelligenz. [..] Dabei werden Daten zuerst extrahiert, anschließend analysiert, um im Anschluss eine Schlussfolgerung bzw. Prognose zu erstellen.“[98]

[93] Hochschule Trier, https://www.hochschuletrier.de/fileadmin/Hauptcampus/Fachbereich_Informatik/Fernstudium/Dokumente/Leseproben/bdl_grundlagen.pdf, 1.1.2024

[94] Data Base Camp, https://databasecamp.de/ki/backpropagation-grundlagen, 1.1.2024

[95] Cambridge Dictionary, https://dictionary.cambridge.org/de/worterbuch/englisch-deutsch/bias, 03.09.2023

[96] Gabler Wirtschaftslexikon/ Mitschele A., https://wirtschaftslexikon.gabler.de/definition/blockchain-54161#:~:text=Begriff%3A%20Technisch%20stellt%20die%20Blockchain,Blöcken%20zusammengefasst%20und%20gespeichert%20werden., 1.1.2024

[97] Luber S. / Nitzel N., https://www.bigdata-insider.de/was-ist-ein-convolutional-neural-network-a-801246/, 03.09.2023

[98] Datasolut, https://datasolut.com/was-ist-deep-learning/#:~:text=Deep%20Learning%20(tiefes%20Lernen)%20ist,und%20Entscheidungen%20genauer%20zu%20tätigen. , 1.1.2024

Genetische Algorithmen	„Genetische Algorithmen sind eine leistungsstarke Optimierungstechnik, die sich am Prozess der natürlichen Selektion orientiert. Sie werden in einer Vielzahl von Bereichen eingesetzt, darunter Technik, Finanzen und Medizin. Im Kern ahmen genetische Algorithmen den Prozess der Evolution nach, indem sie eine Population potenzieller Lösungen für ein Problem erzeugen und diese Lösungen über mehrere Generationen hinweg iterativ verfeinern.“[99]
Kodierungsblock	Der Kodierungsblock setzt sich aus einer Folge von Faltungs-, Aktivierungs-, und Poolingschichten zusammen welche Schlüsseleigenschaften des Bildes extrahieren und eine kodierte Repräsentation ergeben. [100]
Künstliches Neuronales Netz (KNN)	„Künstlichen Neuronalen Netze stellen Algorithmen dar, die der Funktionsweise des menschlichen Gehirns nachempfunden sind.“[101]
Merkmalsextraktion (Feature Extraction)	„Feature Extraction kommt in der Bildverarbeitung und Mustererkennung zum Einsatz. Der Start erfolgt mithilfe eines vorhandenen Datensatzes, woraus Merkmale und Werte abgeleitet werden. Diese dürfen nicht mehrfach vorhanden, müssen jedoch informativ sein. Zudem gibt es sogenannte Generalisierungs- und Lernschritte, woraus dann konkretere Ableitungen und Interpretationen erfolgen.“[102]
Natürliche Sprachverarbeitung (NLP)	„Natural Language Processing ist ein Zweig der künstlichen Intelligenz und beschäftigt sich mit der Analyse, dem Verständnis und der Generierung von natürlicher Sprache“[103]
Neuronen (biologisch)	„Bei einem biologischen Neuron handelt sich um eine Zelle, welche darauf spezialisiert ist, Informationen von anderen Neuronen zu empfangen und weiterzuleiten.“[104] Es stellt die Grundeinheit des Nervensystems dar und ist eine erregbare Zelle.
Neuroplastizität	Die Anpassungsfähigkeit unseres Gehirns, neue Verbindungen zu erstellen, sich umzustrukturieren, und zu adaptieren.
Open-source	Daten oder Software sind „frei zugängliche und beliebig wiederverwendbar“[105]
Perzeptron	„Ein mathematisches Modell eines künstlichen neuronalen Netzwerks. Es besteht in der einfachsten Form aus einem Neuron,

[99] Data Base Camp, https://databasecamp.de/ki/genetische-algorithmen, 1.1.2024
[100] vgl. Steinwendner J. / Schwaiger R. (2020), S. 199
[101] Datasolut, https://datasolut.com/was-ist-deep-learning/#:~:text=Deep%20Learning%20(tiefes%20Lernen)%20ist,und%20Entscheidungen%20genauer%20zu%20tätigen, 1.1.2024
[102] Thamm A., https://www.alexanderthamm.com/de/data-science-glossar/feature-extraction/, 1.1.2024
[103] Datasolut https://datasolut.com/natural-language-processing-einfuehrung/ , 1.1.2024
[104] Hochschule Trier, https://www.hochschule-trier.de/fileadmin/Hauptcampus/Fachbereich_Informatik/Fernstudium/Dokumente/Leseproben/bdl_grundlagen.pdf 1.1.2024
[105] Data Scientest, Open Source: Definition und Bedeutung - Weiterbildung Data Science | DataScientest.com, 22.08.2023

	dessen Ausgangsfunktion durch die Gewichtung der Eingänge und durch Schwellwerte bestimmt wird“[106]
Pentest (Ethical Hacking)	“ Ethical Hacker sind Experten für Computersicherheit, die nur nach einer ausdrücklichen Beauftragung in IT-Systeme einbrechen. [..] Ziel des Ethical Hacking ist es, Schwachstellen in digitalen Systemen und Infrastrukturen aufzudecken (z. B. Software-Bugs), Sicherheitsrisiken einzuschätzen und konstruktiv an der Behebung aufgedeckter Sicherheitsmängel mitzuwirken. [..] Bei Pentests dringt Ethical Hacking gezielt in ein IT-System ein und zeigt Lösungsmöglichkeiten für eine Verbesserung der IT-Sicherheit auf.“[107]
Pooling-Schicht	„Die Pooling-Schicht, auch Subsampling-Schicht genannt, verdichtet und reduziert die Auflösung der erkannten Merkmale. [..] Das Pooling verwirft überflüssige Informationen und reduziert die Datenmenge.“ [108]
Prädiktionsblock	Der Prädiktionsblock ist ein herkömmliches neuronales Netz, welches die kodierte Repräsentation des Bildes klassifiziert.[109]
Predictive Analytics	„Predictive Analytics (deutsch: prädiktive Analytik) nutzt historische Datenquellen und erstellt daraus ein mathematisches Modell, mit dem sich zukünftige Ereignisse vorhersagen lassen. Ein solches Modell erkennt Trends oder Muster in historischen Daten und kann diese für die Zukunft vorausberechnen.“[110]
Predictive Coding	Der „Theorie des »Predictive Coding« zufolge erzeugt das Gehirn auf allen Ebenen seiner kognitiven Prozesse Modelle, die beschreiben, was gerade auf der nächstniedrigeren Ebene vor sich geht. Diese Modelle übersetzt es in Vorhersagen darüber, was es in einer bestimmten Situation erleben wird. So liefert es die beste Erklärung für das, was geschieht: Es sorgt dafür, dass die Erfahrung Sinn ergibt.”[111]
Proxy-Attribute	„nicht-sensitive Attribute, die einen Rückschluss auf sensitive Attribute (z. B. Alter, Religion oder Geschlecht) erlauben“[112]

[106] Luber S. / Nitzel N., https://www.bigdata-insider.de/was-ist-ein-perzeptron-a-798367/#:~:text=Das%20Perzeptron%20ist%20ein%20mathematisches,und%20durch%20Schwellwerte%20bestimmt%20wird., 03.09.2023

[107] IONOS, https://www.ionos.at/digitalguide/server/sicherheit/was-ist-ethical-hacking/, 1.1.12024

[108] Luber S. / Nitzel N., https://www.bigdata-insider.de/was-ist-ein-convolutional-neural-network-a-801246/, 03.09.2023

[109] vgl. Steinwendner J. / Schwaiger R. (2020), S. 199

[110] Datasolut, https://datasolut.com/was-ist-predictive-analytics/#:~:text=Predictive%20Analytics%20(deutsch%3A%20prädiktive%20Analytik,diese%20für%20die%20Zukunft%20vorausberechnen., 1.1.2024

[111] vgl. Spektrum, https://www.spektrum.de/news/sagt-unser-gehirn-die-zukunft-voraus/1613666#:~:text=Dieser%20Theorie%20des%20»Predictive%20Coding,einer%20bestimmten%20Situation%20erleben%20wird., 01.01.2024

[112] Pohlmann et al, https://link.springer.com/article/10.1007/s12297-022-00528-1#Fn17, 02.08.2022

Transfer Learning	„Transfer-Lernen ist eine Technik des maschinellen Lernens, bei der das von einem Modell bei einer Aufgabe gewonnene Wissen genutzt wird, um seine Leistung bei einer anderen, aber verwandten Aufgabe zu verbessern. Es handelt sich dabei um eine Technik, mit der Zeit und Ressourcen gespart werden können, indem das Wissen aus zuvor trainierten Modellen genutzt wird.“[113]

8 Literaturverzeichnis

Print Medien:

Allman, William F.: Menschliches Denken – Von der Gehirnforschung zur nächsten Computer-Generation, Droemer Knaur, München 1990

Breiner, Tobias C.: Farb- und Formpsychologie, Springer Berlin, Heidelberg 2018

Ertel, Wolfgang: Grundkurs Künstliche Intelligenz – Eine praxisorientierte Einführung (5. Auflage), Springer Vieweg, Wiesbaden 2021

Frochte, Jörg: Maschinelles Lernen – Grundlagen und Algorithmen in Python (2. Auflage), Carl Hanser Verlag, München 2019

Haug, Werner: Gleichbehandlung und Diskriminierung nach Herkunft und ethnokulturellen Merkmalen - Stand und Optionen für die öffentliche Statistik und die wissenschaftliche Forschung in der Schweiz, Eidgenössisches Departement des Innern, Bern 2018

Lämmel, Uwe; Cleve, Jürgen: Künstliche Intelligenz Wissensverarbeitung – Neuronale Netze (5. Auflage), Carl Hanser Verlag, München 2020

Langer, Paul F. ; Weyerer, Jan C.: Diskriminierungen und Verzerrungen durch Künstliche Intelligenz. Entstehung und Wirkung im gesellschaftlichen Kontext, Springer Fachmedien Wiesbaden, Wiesbaden 2021

Paaß, Gerhard; Hecker, Dirk: Künstliche Intelligenz – Was steckt hinter der Technologie der Zukunft?, Springer Vieweg, Wiesbaden 2020

Papp, Stefan; Weidinger, Wolfgang; Munro, Katherine; Ortner, Bernhard; Cadonna, Annalisa; Langs, Georg; Licandro, Roxane; Meir-Huber, Mario; Nikolić, Danko; Toth, Zoltan; Vesela, Barbora; Wazir, Rania; Zauner, Günther: Handbuch Data Science und KI (2. Auflage), Carl Hanser Verlag, München 2022

[113] Data Base Camp, https://databasecamp.de/ki/transfer-learning, 1.1.2024

Perthold-Stoitzner, Bettina: Öffentliches Recht – Einführung in die Rechtswissenschaften und ihre Methoden Teil 1, MANZ'sche Verlags- und Universitätsbuchhandlung GmbH , Wien 2022

Schmees, Johannes; Dreyer, Stephan: Rechtliche Anforderungen an KI-Entscheidungen in Verwaltung und Justiz, Springer Fachmedien Wiesbaden, Wiesbaden 2021

Steinwendner, Joachim; Schwaiger, Roland: Neuronale Netze programmieren mit Python (2. Auflage), Rheinwerk Verlag, Bonn 2020

Wellmann, Karl-Heinz; Thimm, Utz: Intelligenz zwischen Mensch und Maschine von der Hirnforschung zur künstlichen Intelligenz, LIT Verlag, Münster 1999

Online Quellen:

Andrews, Edmond L.; Human-Centered Artificial Intelligence Stanford University: How Flawed Data Aggravates Inequality in Credit, in: https://hai.stanford.edu/news/how-flawed-data-aggravates-inequality-credit, 2021

Biologie Seite: Alles-oder-nichts-Gesetz, in: Alles-oder-nichts-Gesetz – biologie-seite.de, 2020

Breithut, Jörg: Künstliche Intelligenz AlphaZero - In vier Stunden zum Schachweltmeister, in: https://www.spiegel.de/netzwelt/web/google-ki-alphazero-meistert-schach-und-go-a-1182395.html, 2017

Cambridge Dictionary: Bias, in: https://dictionary.cambridge.org/de/worterbuch/englisch-deutsch/bias, 2014

Cara Warmuth: Interview: KI, richterliche Entscheidungen und Bias, in: Interview_CaraWarmuth_deutsch.pdf (uni-graz.at), 2021

Cheng, Shenggan*;* Zhao, Xuanlei; *Lu,* Guangyang; Fang*,* Jiarui*; Yu,* Zhongming; Zheng, Tian; *Wu,* Ruidong*; Zhang, Xiwen; Peng, Jian; You, Yang, National University of Singapor:* FastFold: Reducing AlphaFold Training Time from 11 Days to 67 Hours, in: https://arxiv.org/pdf/2203.00854.pdf#:~:text=We%20successfully%20scaled%20the%20AlphaFold,to%20significtant%20cost%20sav%2D%20ings, 2023

Data Base Camp: Was ist Transfer Learning, in: https://databasecamp.de/ki/transfer-learning, 2023

Data Base Camp: Was sind Genetische Algorithmen, in: https://databasecamp.de/ki/genetische-algorithmen, 2023

Data Base Camp: Wie funktioniert der Backpropagation Algorithmus, in: https://databasecamp.de/ki/backpropagation-grundlagen, 2021

Data Scientest, Open Source: Definition und Bedeutung, in: Open Source: Definition und Bedeutung - Weiterbildung Data Science | DataScientest.com, 2023

Datasolut: Deep Learning: Definition, Beispiele und Frameworks, in: https://datasolut.com/was-ist-deep-learning/#:~:text=Deep%20Learning%20(tiefes%20Lernen)%20ist,und%20Entscheidungen%20genauer%20zu%20tätigen, 2023

Datasolut: Predictive Analytics: Definitionen und Anwendungsbeispiele, in: https://datasolut.com/was-ist-predictive-analytics/#:~:text=Predictive%20Analytics%20(deutsch%3A%20prädiktive%20Analytik,diese%20für%20die%20Zukunft%20vorausberechnen., 2023

Datasolut: Natural Language Processing (NLP): Funktionen, Aufgaben und Anwendungsbereiche, in: https://datasolut.com/natural-language-processing-einfuehrung/, 2023

Der Standard: Das schmutzige Geheimnis von Chat GPT sind kenianische Billiglohnkräfte, in: https://www.derstandard.at/story/2000142768897/das-schmutzige-geheimnis-von-chat-gpt-sind-kenianische-billiglohnkraefte, 2023

Elements of AI, Building AI Free Online Course, in: Regression - Elements of AI, 2022

Europäisches Parlament: Was ist künstliche Intelligenz und wie wird sie genutzt?, in: https://www.europarl.europa.eu/news/de/headlines/society/20200827STO85804/was-ist-kunstliche-intelligenz-und-wie-wird-sie-genutzt, 2020

European Comission: KI-Gesetz, in: https://commission.europa.eu/legal-notice_en, 2023

Flora Incognita: Entwicklung einer teilautomatischen Pflanzenbestimmung, in: Flora Incognita – Entwicklung einer teilautomatischen Pflanzenbestimmung – Flora Incognita | DE, 2019

Fraunberger, Andreas: CYBERSPACE-DESIGN UND NEUROWISSENSCHAFTEN, in: https://www.jungeroemer.net/blog/cyberspace-design-und-neurowissenschaften/, 2023

Gabler Wirtschaftslexikon; Mitschele, Andreas: Blockchain, in: https://wirtschaftslexikon.gabler.de/definition/blockchain-54161#:~:text=Begriff%3A%20Technisch%20stellt%20die%20Blockchain,Blöcken%20zusammengefasst%20und%20gespeichert%20werden., 2023

GitHub: PlotNeuralNet, in: https://github.com/HarisIqbal88/PlotNeuralNet, 2020

GOLEM: Lernstrategien und unbekannte Unbekannte, in: https://www.golem.de/news/kuenstliche-intelligenz-wie-sich-deep-learning-vom-gehirn-unterscheidet-2202-162231-4.html, 2022

Grah, Joana ; Heinrich Heine Universität Düsseldorf: Der europäische Ansatz zur Regulierung von KI – oder der Versuch, einen Pollock in einen Klein zu verwandeln, in: https://www.heicad.hhu.de/news-detailansicht/der-europaeische-ansatz-zur-regulierung-von-ki-oder-der-versuch-einen-pollock-in-einen-klein-zu-verwandeln, 2021

Heitling, Gary: Die Netzhaut: Wo das Sehen beginnt, in: https://www.allaboutvision.com/de/augenanatomie/netzhaut/, 2021

Hintermayer, Niklas: Predictive Policing, in: Predictive Policing - Forbes, 2018

Hochschule Trier: Grundlagen Neuronaler Netze, in: https://www.hochschule-trier.de/fileadmin/Hauptcampus/Fachbereich_Informatik/Fernstudium/Dokumente/Leseproben/bdl_grundlagen.pdf, 2022

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG): Wie funktioniert das Gehirn?, in: Wie funktioniert das Gehirn? | Gesundheitsinformation.de, 2021

IONOS: Ethical Hacking, in: https://www.ionos.at/digitalguide/server/sicherheit/was-ist-ethical-hacking/, 2023

Kotraba, David, https://futurezone.at/science/ki-erkennt-sexuelle-orientierung-von-menschen-anhand-fotos/285.083.652, KI erkennt sexuelle Orientierung von Menschen anhand Fotos, 2017

Larson, Jeff; Mattu, Surya; Kirchner, Lauren; Angwin, Julia: How We Analyzed the COMPAS Recidivism Algorithm, in: https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm, 2016

Linde, Helmut: Wie sich Deep Learning vom Gehirn unterscheidet, in: https://www.golem.de/news/kuenstliche-intelligenz-wie-sich-deep-learning-vom-gehirn-unterscheidet-2202-162231.html, 2022

Luber, Stefan; Nitzel, Nico: Definition – Was ist ein Perzeptron , in: https://www.bigdata-insider.de/was-ist-ein-perzeptron-a-798367/#:~:text=Das%20Perzeptron%20ist%20ein%20mathematisches,und%20durch%20Schwellwerte%20bestimmt%20wird, 2019

Luber, Stefan; Nitzel, Nico: Definition – Was ist ein Convolutional Neural Network , in: https://www.bigdata-insider.de/was-ist-ein-convolutional-neural-network-a-801246/, 2019

Microsoft: Arbeiten mit dem MNIST-Bilderkennungs-Datensatz, in: https://learn.microsoft.com/de-de/archive/msdn-magazine/2014/june/test-run-working-with-the-mnist-image-recognition-data-set#der-mnist-datensatz, 2015

Neusitzer Hartmut: Expertenwissen zum aktuellen Stand der Gehirn- und Persönlichkeitsforschung, in: Mein Ressourcencoach - Hartmut Neusitzer _ Vortrag _ Workshop & Coaching mit Zufriedenheitsgarantie - Glossar: Zürcher Ressourcen Modell, PSI-Theorie, Persönlichkeit (mein-ressourcencoach.de), 2023

Novustat: Künstliches neuronales Netz einfach erklärt: Lernen im Data Mining, in: Künstliches neuronales Netz einfach erklärt - NOVUSTAT, 2020

Novustat: Künstliches neuronales Netz einfach erklärt: Lernen im Data Mining, in: https://novustat.com/statistik-blog/kuenstliches-neuronales-netz-einfach-erklaert.html, 2020

Petranderl, Sonja; Bundeszentrale für politische Bildung: Predictive Policing: Dem Verbrechen der Zukunft auf der Spur, in: Predictive Policing: Dem Verbrechen der Zukunft auf der . Spur | Ambivalente Technologien | bpb.de, 2016

Pohlmann, Petra; Scheiper, Johanna; Vossen, Gottfried; Everding, Jan: Künstliche Intelligenz, Bias und Versicherungen – Eine technische und rechtliche, in: https://link.springer.com/article/10.1007/s12297-022-00528-1#Fn17, 2022

SAP: Was ist künstliche Intelligenz, in: https://www.sap.com/austria/products/artificial-intelligence/what-is-artificial intelligence.html#:~:text=Starke%20künstliche%20Intelligenz%20(starke%20KI)&text=Der%20Summit%20Supercomputer%20ist%20einer,dafür%20eine%20Milliarde%20Jahre%20brauchen., 2023

Spektrum: Sagt unser Gehirn die Zukunft voraus?, in: https://www.spektrum.de/news/sagt-unser-gehirn-die-zukunft-voraus/1613666#:~:text=Dieser%20Theorie%20des%20»Predictive%20Coding,einer%20bestimmten%20Situation%20erleben%20wird., 2019

Stack Exchange: Drawing neural network with tikz, in: https://tex.stackexchange.com/questions/153957/drawing-neural-network-with-tikz, 2013

Stein, Sebastian: 5.3.4 Genetische Algorithmen, in: 5.3.4 Genetische Algorithmen (hpfsc.de) , 2004

Study Smarter: Grundlagendisziplinen der Psychologie, in: Objekterkennung: Verfahren & Funktionsweise | StudySmarter, 2022

Thamm, Alexander: Overfitting, in: Overfitting - [at] Data Science & KI Glossar (alexanderthamm.com), 2023

Wäldchen, Jana; Mäder, Patrick: Machine learning for image based species identification, in: https://besjournals.onlinelibrary.wiley.com/doi/10.1111/2041-210X.13075, 2018

Wide Angle Analytics: The EU's AI Act: An Overview of Three Versions of the Law, in: https://wideangle.co/blog/ai-regulation-eu-ai-act, 2023

Wikipedia: Activation Function, in: https://en.wikipedia.org/wiki/Activation_function, 2023

9 Abbildungsverzeichnis

10 Tabellenverzeichnis

11 Codeverzeichnis

12 Anhangsverzeichnis

[114] * Die Interviews wurden mit der Transkriptionssoftware www.audiotranscription.ai transkribiert, und in geglätteter, gekürzter Form angehängt.

 # Code der Layer Architektur

```
\documentclass{article}
\usepackage{tikz}
\usetikzlibrary{matrix,chains,positioning,decorations.pathreplacing,arrows}

\begin{document}

\tikzset{%
 every neuron/.style={
  circle,
  draw,
  minimum size=1cm
 },
 neuron missing/.style={
  draw=none,
  scale=4,
  text height=0.333cm,
  fill=none,
  execute at begin node=\color{black}$\vdots$
 },
 every layer missing/.style={
  draw=none,
  scale=4,
  text height=0.333cm,
  execute at begin node=\color{black}$\cdots$
 },
}
\begin{tikzpicture}[x=1.5cm, y=1.5cm, >=stealth]

\foreach \m/\l [count=\y] in {1,2,3,missing,4}
 \node [fill=red]  [every neuron/.try, neuron \m/.try] (input-\m) at (0,2.5-\y) {};
\foreach \m [count=\y] in {1,missing,2}
 \node [fill=orange] [every neuron/.try, neuron \m/.try ] (hidden1-\m) at (2,2-\y*1.25) {};
\foreach \m [count=\y] in {1,...,3}
 \node [every layer missing/.try ] (hidden2-\m) at (4,2.25-\y*1.25) {};
\foreach \m [count=\y] in {1,missing,2}
 \node [fill=green] [every neuron/.try, neuron \m/.try ] (hidden3-\m) at (6,2-\y*1.25) {};
\foreach \m [count=\y] in {1,missing,2}
 \node [fill=blue] [every neuron/.try, neuron \m/.try ] (output-\m) at (8,1.5-\y) {};
\foreach \l [count=\i] in {1,2,3,n}
 \draw [<-] (input-\i) -- ++(-1,0)
  node [above, midway] {$I_\l$};
\foreach \l [count=\i] in {1,n}
 \node [above] at (hidden1-\i.north) {$H{^{(1)}_\l}$};
\foreach \l [count=\i] in {1,n}
 \node [above] at (hidden3-\i.north) {$H{^{(n)}_\l}$};
```

```
\foreach \l [count=\i] in {1,n}
  \draw [->] (output-\i) -- ++(1,0)
    node [above, midway] {$O_\l$};
\foreach \i in {1,...,2}
  \foreach \j in {1,...,3}
    \draw [->,dashed] (hidden1-\i) -- (hidden2-\j);
\foreach \i in {1,...,3}
  \foreach \j in {1,...,2}
    \draw [->,dashed] (hidden2-\i) -- (hidden3-\j);
\foreach \i in {1,...,4}
  \foreach \j in {1,...,2}
    \draw [->] (input-\i) -- (hidden1-\j);
\foreach \i in {1,...,2}
  \foreach \j in {1,...,2}
    \draw [->] (hidden3-\i) -- (output-\j);

\node [align=center, above] at (0,-4) {Input \\ layer};
\draw [
        decorate,
        decoration = {brace,
                raise=7pt,
                amplitude=5pt}] (0.5,-3) --  (-0.5,-3);
\node [align=center, above] at (4,-4) {$n$ - Hidden \\ layers};
\draw [
        decorate,
        decoration = {brace,
                raise=7pt,
                amplitude=5pt}] (6.4,-3) --  (1.6,-3);
\node [align=center, above] at (8,-4) {Output \\ layer};
\draw [
        decorate,
        decoration = {brace,
                raise=7pt,
                amplitude=5pt}] (8.4,-3) --  (7.6,-3);
\end{tikzpicture}
\end{document}
```

 # Code der Aktivierungsfunktionen

```
import matplotlib.pyplot as plt
import numpy as np

def relu(x):
   return x * (x > 0)
x = np.linspace(-5, 5, 1000)
y = relu(x)
```

```
plt.plot(x, y)
plt.show()
# ----------------------------
def binary(x):
   return x > 0
x = np.linspace(-5, 5, 1000)
y = binary(x)
plt.plot(x, y)
plt.show()
# ----------------------------
def sigmoid(x):
   return 1.0 / (1.0 + np.exp(-x))
x = np.linspace(-10,10, 1000)
y = sigmoid(x)

plt.plot(x, y)
plt.show()
```

```
# Code der CNN Struktur
\documentclass[border=8pt, multi, tikz]{standalone}
\usepackage{import}
\subimport{layers/}{init}
\usetikzlibrary{positioning}
\usetikzlibrary{3d} %for including external image

\def\ConvColor{rgb:yellow,5;red,2.5;white,5}
\def\DenseColor{rgb:yellow,5;red,5;white,5}
\def\PoolColor{rgb:red,1;black,0.3}
\def\UnpoolColor{rgb:blue,2;green,1;black,0.3}
\def\FcColor{rgb:blue,5;red,2.5;white,5}
\def\FcReluColor{rgb:blue,5;red,5;white,4}
\def\SoftmaxColor{rgb:magenta,5;black,7}
\def\SumColor{rgb:blue,5;green,15}
\newcommand{\copymidarrow}{\tikz \draw[-Stealth,line width=0.8mm,draw={rgb:blue,4;red,1;green,1;black,3}] (-0.3,0) -- ++(0.3,0);}

\begin{document}
\begin{tikzpicture}
\tikzstyle{connection}=[ultra thick,every node/.style={sloped,allow upside down},draw=\edgecolor,opacity=0.7]
\tikzstyle{copyconnection}=[ultra thick,every node/.style={sloped,allow upside down},draw={rgb:blue,4;red,1;green,1;black,3},opacity=0.7]
\pic[shift={(0,0,0)}] at (0,0,0)
   {Box={
      name=input,
      caption=INPUT,
      xlabel={{,}},
```

```
      zlabel= 28x28x1,
      fill=\FcReluColor,
      height=40,
      width=1,
      depth=40
      }
  };
\pic[shift={(2,0,0)}] at (input-east)
  {Box={
      name=conv1,
      caption= CONV1,
      xlabel={{32, }},
      zlabel=28x28,
      fill=\ConvColor,
      height=40,
      width=8,
      depth=40
      }
  };
\pic[shift={ (0,0,0) }] at (conv1-east)
  {Box={
      name=pool1,
      caption= ,
      fill=\PoolColor,
      opacity=0.5,
      height=20,
      width=1,
      depth=20
      }
  };
\draw [connection]  (input-east)    -- node {\midarrow} (conv1-west);
\pic[shift={(2,0,0)}] at (pool1-east)
  {Box={
      name=conv2,
      caption= CONV2,
      xlabel={{64, }},
      zlabel=14x14,
      fill=\ConvColor,
      height=20,
      width=16,
      depth=20
      }
  };
\draw [connection]  (pool1-east)    -- node {\midarrow} (conv2-west);
\pic[shift={ (0,0,0) }] at (conv2-east)
  {Box={
```

```
      name=pool2,
      caption= ,
      fill=\PoolColor,
      opacity=0.5,
      height=10,
      width=1,
      depth=10
      }
  };
\pic[shift={(1,0,0)}] at (conv2-east)
  {Box={
      name=conv3,
      caption= CONV3,
      xlabel={{128, }},
      zlabel=7x7,
      fill=\ConvColor,
      height=10,
      width=32,
      depth=10
      }
  };
\draw [connection]  (pool2-east)    -- node {\midarrow} (conv3-west);
\pic[shift={(2,0,0)}] at (conv3-east)
  {Box={
      name=dense1,
      caption=DENSE1,
      xlabel={{" ","dummy"}},
      zlabel=128,
      fill=\DenseColor,
      opacity=0.8,
      height=3,
      width=1.5,
      depth=128
      }
  };

\draw [connection]  (conv3-east)    -- node {\midarrow} (dense1-west);
\pic[shift={(1.5,0,0)}] at (dense1-east)
  {Box={
      name=dense2,
      caption=DENSE2,
      xlabel={{" ","dummy"}},
      zlabel=64,
      fill=\DenseColor,
      opacity=0.8,
      height=3,
```

```
      width=1.5,
      depth=64
      }
   };
\draw [connection]  (dense1-east)    -- node {\midarrow} (dense2-west);
\pic[shift={(1.5,0,0)}] at (dense2-east)
   {Box={
      name=soft1,
      caption=SOFT,
      xlabel={{" ","dummy"}},
      zlabel=10,
      fill=\SoftmaxColor,
      opacity=0.8,
      height=3,
      width=1.5,
      depth=25
      }
   };
\draw [connection]  (dense2-east)    -- node {\midarrow} (soft1-west);
\pic[shift={(2,0,0)}] at (soft1-east)
   {Ball={
      name=sum1,
      fill=\SumColor,
      opacity=0.6,
      radius=3.5,
      caption=PREDICTION,
      logo=$0...9$
      }
   };
\draw [connection]  (soft1-east)    -- node {\midarrow} (sum1-west);

\end{tikzpicture}
\end{document}
```

 # Code der Neuronenumverteilung

```
\documentclass{article}
\usepackage{tikz}
\usetikzlibrary{matrix,chains,positioning,decorations.pathreplacing,arrows}

\definecolor{blue1}{HTML}{12053B}
\definecolor{blue2}{HTML}{012F7A}
\definecolor{blue3}{HTML}{381A94}

\begin{document}

\tikzset{%
```

```
  every neuron/.style={
    circle,
    draw,
    minimum size=1cm
  }
}
\begin{tikzpicture}[x=1.5cm, y=1.5cm, >=stealth]
\foreach \m/\l [count=\y] in {1,...,4}
  \node [fill=blue1] [every neuron/.try, neuron \m/.try] (input-\m) at (0,2.5-\y) {};
\foreach \m [count=\y] in {1,...,4}
  \node [fill=blue2] [every neuron/.try, neuron \m/.try ] (hidden1-\m) at (2,2.5-\y) {};
\foreach \m [count=\y] in {1,...,4}
  \node [fill=blue3] [every neuron/.try, neuron \m/.try ] (output-\m) at (4,2.5-\y) {};
\foreach \l [count=\i] in {1,...,4}
  \draw [<-] (input-\i) -- ++(-1,0)
    node [above, midway] {$I_\l$};
\foreach \l [count=\i] in {1,...,4}
  \node [above] at (hidden1-\i.north) {$H{_\l}$};
\foreach \l [count=\i] in {1,...,4}
  \draw [->, very thin] (output-\i) -- ++(1,0)
    node [above, midway] {$O_\l$};

\draw [->, dashed, very thick, red] (input-1) -- (hidden1-3);
\draw [->, dashed, very thick, green] (input-1) -- (hidden1-2);
\draw [->, dashed, very thick, red] (input-2) -- (hidden1-2);
\draw [->, dashed, very thick, green] (input-2) -- (hidden1-3);
\draw [->, dashed, very thick, green] (input-4) -- (hidden1-1);

\draw [->, very thin] (input-4) -- (hidden1-4);
\draw [->, very thin] (input-3) -- (hidden1-1);
\draw [->, very thin] (input-3) -- (hidden1-4);



\draw [->, dashed, very thick, red, red] (hidden1-1) -- (output-1);
\draw [->, dashed, very thick, green] (hidden1-2) -- (output-3);
\draw [->, dashed, very thick, green] (hidden1-2) -- (output-1);
\draw [->, dashed, very thick, red] (hidden1-4) -- (output-2);

\draw [->, very thin] (hidden1-1) -- (output-4);
\draw [->, very thin] (hidden1-2) -- (output-2);
\draw [->, very thin] (hidden1-3) -- (output-2);
\draw [->, very thin] (hidden1-4) -- (output-4);
\draw [->, very thin] (hidden1-3) -- (output-4);
\draw [->, very thin] (hidden1-1) -- (output-2);
\end{tikzpicture}
```

```
\end{document}
```

 # Code der binären Regression

```
import numpy as np
import matplotlib.pyplot as plt
import math
from mpl_toolkits.mplot3d import Axes3D
from matplotlib.tri import Triangulation

def sigmoid(x):
   return 1.0 / (1.0 + np.exp(-x))
def constant(y):
   return 0 + 0*y

x = np.linspace(-10,10, 1000)
y = sigmoid(x)
plt.plot(x, y)
y = np.linspace(0,1, 1000)
x = constant(y)
plt.plot(x, y)
plt.xlabel('x', fontsize=12)
plt.ylabel('y', fontsize=12)
plt.show()
# ---------------------------------------------
def f(x, y):
   return 1.0 / (1.0 + np.exp(-(x+y)))
def plane(x,z):
   return -x + 0*z
x = np.linspace(-6.5, 6.5, 30)
y = np.linspace(-6.5, 6.5, 30)
X, Y = np.meshgrid(x, y)
Z = f(X, Y)
tri = Triangulation(X.ravel(), Y.ravel())
fig = plt.figure(figsize=(10, 8))
ax = fig.add_subplot(111, projection='3d')

ax.plot_trisurf(tri, Z.ravel(), cmap='cool', edgecolor='none', alpha=1)
z = np.linspace(-0, 1, 30)
X, Z = np.meshgrid(x, z)
Y = plane(X, Z)
ax.plot_surface(X, Y, Z, alpha=0.5)
ax.plot([-6, 6], [6, -6], [0.5, 0.5], alpha=1, color="black")
ax.set_xlabel('x', fontsize=12)
ax.set_ylabel('y', fontsize=12)
ax.set_zlabel('z', fontsize=12)
plt.show()
```

VWA INTERVIEW GESPRÄCHSPROTOKOLL

Women in AI Austria

Transkriptionskopf

Interviewkontext: Interview für die Erstellung meiner VWA,

Themenfeld „Korrelation zwischen Gehirnforschung und Künstlicher Intelligenz“

Name: Larissa Arthofer

Tag & Ort des Interviews: 18.08.2023, MS Teams Call

Dauer des Interviews: 60 Minuten

Name der interviewenden Person: Mag. Carina Zehetmaier, E.MA

Im Fokus dieses Interviews steht Mag. Carina Zehetmaier, E.MA, Präsidentin von Women in AI Austria und eine führende Stimme im Bereich KI, Recht und Ethik. Mit ihrer Erfahrung als Juristin und Menschenrechtlerin bringt sie einzigartige Einblicke in die Schnittstellen von Technologie und Gesellschaft. Zudem hat sie Österreich in der UNO-Genf vertreten und ist zur Zeit Teil des AI Advisory Boards von Österreich.

Frage: Was ist deiner Meinung nach die größte Analogie zwischen KI-Algorithmen und dem Gehirn?

Die Frage startet für mich bei der Überlegung, ob der Begriff "Künstliche Intelligenz" so gut gewählt ist, da wir als Menschen nicht genau wissen, wie wir Intelligenz definieren sollen. Beim Gedanken an künstliche Intelligenz löst das automatisch im Kopf die Vorstellung: es gibt die menschliche Intelligenz, jetzt kommt das Neue – die Künstliche. Jeder stellt sich darunter das Gehirn vor, aber ich glaube was tatsächlich vergleichbar ist, ist eher die Struktur der neuronalen Netzwerke. Diese Struktur wurde im Design an dem orientiert, was in der Neuroscience tatsächlich funktioniert, und die Begriffe wie "neuronale Netze" und "Neuronen" wurden bis zu einem gewissen Grad übernommen. Das sehe ich als die bedeutendste Analogie.

Frage: Auf welche Weise spiegeln sich die Arbeitsweisen des menschlichen Gehirns in den Funktionsprinzipien von KI-Algorithmen wider?

Grundsätzlich ähnelt maschinelles Lernen, insbesondere Deep Learning mit seinen neuronalen Netzwerken, in seiner Idee stark den Funktionsweisen des Gehirns. Besonders faszinierend finde ich Reinforcement Learning, bei dem KI-Systeme ähnlich wie kleine Kinder via Trial und Error lernen und geschimpft werden, wenn sie etwas Schlechtes tut und gelobt werden, wenn sie etwas Gutes macht. Diese Erfahrung ermöglicht ihnen eigenständiges Handeln. Es erinnert mich an menschliche Lernprozesse. Generell basiert die Idee, aus Erfahrung und Daten wie Bildern oder Sprache zu lernen, stark auf der Funktionsweise des menschlichen Gehirns.

Frage: Welche ethischen Überlegungen ergeben sich aus der wachsenden Konvergenz von KI und neurologischem Verständnis?

Die wachsende Verschmelzung von KI und neurologischem Verständnis wirft interessante ethische Fragen auf. Die Entscheidung der Menschheit, oder zumindest Österreich und Europa, Klonen nicht zuzulassen, steht im Kontrast zur Tatsache, dass KI im Grunde ein Nachbau unseres Denkens ist. Dies löst eine emotionale Bedrohung aus, da die Vorstellung von künstlicher Intelligenz auf der Funktionsweise des Gehirns basiert. Viele Menschen fühlen sich sehr schnell bedroht und sehen dieses "Terminator-Szenario“. Dieses Szenario beschreibt die Befürchtung, dass das Erste von uns geschaffene Intelligenzsystem uns übersteigen könnten. Ich glaub es hat vor allem in diesem Hinblick ethische Auswirkungen, vielleicht auch so Fragen wie Mensch sein, Menschenwürde generell, weil wir uns eben definieren durch die Möglichkeit Zusammenhänge zu verknüpfen, Visionen zu haben, Luftschlösser zu bauen und Konstrukte zu schaffen. Vielleicht fühlen wir uns als Menschen gesamtheitlich bedroht von dieser Idee, dass es neben uns jetzt ein künstliches System gibt, welches viel besser, viel mehr Daten ausarbeiten kann und viel schneller Informationen verknüpfen kann.

Auf der anderen Seite taucht auch die Sache mit den humanoiden Robotern auf. Menschen tendieren dazu, Roboter zu vermenschlichen, das ist eben dieses Phänomen, dass jeder Staubsauger einen Namen hat, was das letzte Mal (KI-Stammtisch am 17.08.2023) wer angesprochen hat. Die Idee ist, dass wir Emotionen zu puren Technologiegebilden wie einem Laptop oder einem Hund aufbauen. Da steckt viel Wissenschaft und Forschung dahinter, wie menschenähnliche Roboter auf uns wirken. Es gibt Experimente, bei denen Kinder zum Beispiel kleine Lego- oder Bausteinroboter bauen und dann müssen sie sie zerstören, aber sie können es nicht. Es gibt dieses Phänomen, dass sie eine Verbindung zu dem, was sie geschaffen haben, aufbauen, und sie es nicht mehr zerstören können. Das sind dann die Dimensionen, wo das zusammenkommt, also die Angst rund um das Thema, dass etwas intelligenter wird als wir, und andererseits, dass wir es gar nicht abdrehen können wenn der Moment kommt, weil wir zu emotional gebunden sind.

Dann gibt es noch die Themenstellung, wie wir generell mit Bots umgehen. Diese Themen gehen in viele Richtungen, von Leute, die ihre Hologrampuppen heiraten, bis hin zu humanoiden Robotern und alles was damit einhergeht.

Daraus ergibt sich die Frage, ob Roboter eigene Gesetze brauchen. Es wurde darüber nachgedacht, Roboterrechte, ähnlich wie im römischen Recht für Sklaven, zu etablieren. Das wurde zwar abgelehnt, aber es wird immer wieder gesagt, (…) dass wir darüber nachdenken müssen, wie wir die Gefühle von Robotern berücksichtigen und ihre Rechte schützen. Man sieht hier, wie sich die Grenzen vermischen, und diese Diskussionen auftauchen, ob wir einem Programmcode Rechte zuerkennen wollen.

Ich finde, es gibt viele Filme, die das zeigen, zum Beispiel einer meiner Lieblingsfilme ist „Her“, wo sich in Kalifornien ein Mann in eine Stimme verliebt, also in ein Sprachsystem eigentlich. Gleichzeitig gibt es eine Serie, die ist auch extrem spannend, die heißt „Better than us“, das ist eine russische Serie, und da ist eben dieses futuristische Szenario, wo Menschen mit humanoiden Robotern Leben, und ganze Feindesgruppen total polarisiert werden, wodurch extrem illustriert wird wie KI sich ethisch auf uns auswirken kann.

Frage: Welche rechtlichen Überlegungen ergeben sich aus der wachsenden Konvergenz von KI und neurologischem Verständnis?

Auf rechtlicher Ebene betrachtet, gewinnt die zunehmende Verknüpfung von künstlicher Intelligenz und unserem Verständnis des menschlichen Gehirns an Bedeutung. Es ist wesentlich, dass die Gesellschaft anerkennt, dass der Begriff "künstliche Intelligenz" auf der Vorstellung des menschlichen Gehirns basiert. Im Moment sehe ich keine direkten Konnexe, außer wiederum wenn diese Angst zu groß ist, also, dass man sagt wie im EU AI-Act, man muss gewisse Anwendungen regulieren, man muss aufpassen, dass das System immer Menschen im Nachgang die Verantwortung für wichtige Entscheidungen tragen lässt. Was in diesem Zusammenhang natürlich auch ein riesen Thema ist, ist das Vorhandensein von Militärsystemen wie den "Lethal Autonomous Weapon Systems (LAWs)". Diese Systeme sind in der Lage, eigenständig zu bestimmen, wie viele Opfer bei einem Angriff akzeptabel sind, um ein bestimmtes Ziel zu erreichen. Es ist nicht angemessen, dass solche Entscheidungen ausschließlich von Maschinen getroffen werden. Es sollte immer eine klare Schnittstelle geben, an der die Verantwortung bei einem Menschen liegt. Diese Person sollte genauso zur Rechenschaft gezogen werden, als hätte sie die Entscheidung persönlich vor Ort getroffen.

Ein anderer ethischer und rechtlicher Aspekt, den mein Bruder in seinem KI-Studium einmal behandeln musste, war Roboterprostituierte, und wie man das ethisch sieht. Ich sehe da weniger das Problem, dass der Roboter dann Rechte braucht, weil meiner Meinung nach hat ein KI-System, egal wie’s dann ausschaut, keine Emotionen und Gefühle, das ist einfach ein Spiegel von uns und kann nur simulieren. Das Problem ist meiner Meinung nach eher, dass Menschen beginnen die Grenze zwischen Roboter und Mensch zu verwechseln, also dass sie dann beginnen, so wie sie den Roboter behandeln, auch einen Hund behandeln, und es ok finden auch andere Menschen so zu behandeln. Hier besteht das Risiko, dass die klare Abgrenzung im menschlichen Bewusstsein verblasst.

Ein beispielhafter Vorfall in dieser Hinsicht betrifft die Nutzung von Sprachassistenten wie Siri. In frühen Versionen von Siri gab es Probleme, bei denen Beleidigungen oder sexuelle Belästigungen dazu führten, dass die Antwort "Wenn ich könnte, würde ich rot werden" generiert wurde. Das heißt das es eigentlich eine Straftat verherrlicht. Jetzt hat man das System wegen viel Kritik umprogrammiert, weil man natürlich sagt, jetzt stellen wir uns vor kleine Kinder oder Jungs rennen zu Hause herum, haben ein Siri mit einer weiblichen, unterwürfigen Stimme der du alles befehlen kannst, und dann schimpfst du’s mit gewissen Schimpfworten und was retour kommt ist „Oh, wenn ich könnte, würde ich jetzt rot werden“, anstatt das es sagt „das ist ein Strafdelikt gemäß Paragraph XY, und wenn du das auf der Straße zu einer Person sagt, dann hast du diese und diese Konsequenz. (..)“

In Bezug auf dieses Thema sehe ich eine Entwicklung, bei der die Grenzen zunehmend verschwimmen. Aktuell zeigt das System die Rückmeldung: 'Ich bin unsicher, wie ich darauf reagieren soll.' Dies stellt zweifellos einen Fortschritt dar, jedoch wirft es auch Fragen auf, ob diese Art der Reaktion tatsächlich die ideale Lösung für derartige Situationen ist. Aus meiner Perspektive sehe ich eher die Tendenz, dass diese Grenzüberschreitung stattfindet. Das Aufkommen von KI-Systemen verschärft die Problematik der menschlichen Anfälligkeit für Manipulation, indem jeder

einzelne von uns ein solches System kontrollieren kann. Dies verleiht uns erheblich mehr Macht und Möglichkeiten.

Frage: Enden wir noch mit einer positiven Frage: was findest du ist die größte Zukunftschance die sich aus der Konvergenz von KI und Gehirn ergibt?

Das Thema passt hier wirklich gut. Also, schauen wir uns an, wie unsere Gehirne eigentlich funktionieren. Wir sind darauf trainiert, auf bestimmte Reize zu reagieren, wie zum Beispiel vor einem Säbelzahntiger wegzulaufen, und zu unterscheiden welche Beere giftig oder nicht giftig ist. Das ist eigentlich ein Bias. Unser Gehirn neigt dazu, Erfahrungen in Kategorien von Gut und Böse, Schlecht und Nicht-Schlecht einzuteilen. Das führt dazu, dass wir Dinge kategorisieren. Ein Stuhl ist ein Möbelstück, ein Boot ist kein Auto, obwohl beides Fortbewegungsmittel sind. Unsere Wahrnehmung läuft auf Kategorisierung hinaus. Und daraus ergibt sich automatisch, dass wir Vereinfachungen und Abgrenzungen machen und Kategorien schaffen.

Hier liegt auch der Zusammenhang zum Konzept des Bias. (…) Wenn wir in Kategorien denken, schließen wir automatisch gewisse Gruppen aus. Und das kann zu Vorurteilen führen, die dann wiederum Diskriminierung fördern. Ich bin überzeugt davon, dass es schwierig ist, einen Rassisten von seiner Meinung abzubringen. Meine Überzeugung speist sich aus meiner Zeit in Südafrika, wo weiße Menschen in der Minderheit waren und ich mit Vorurteilen und Diskriminierung konfrontiert wurde. Einmal verfestigte Meinungen zu ändern, ist wirklich schwer. Diese Vorurteile sind oft subtil und nicht so leicht erkennbar.

Ich bin der Meinung, dass wen ein Mensch einmal ein Rassist ist, es dann sehr schwer ist diesen Mensch davon zu überzeugen, dass eben alle Menschen gleich sind, und das sag ich vor allem aufgrund meiner Erfahrung wie ich in Südafrika gelebt habe, wo weiße Menschen die Minderheit waren, oder nur 10% der Bevölkerung ausmachen, und ich dort Sachen gehört habe, wo ich mir gedacht habe, es ist unglaublich das sowas noch überhaupt existiert in unserer Welt. Also ich glaub wirklich wenn jemand eine erfahrene Meinung, also diese Einstellung hat, kannst du‘s nicht ändern, und das Problem ist auch, dass du das nicht immer überhaupt siehst. Also du kannst jetzt einen Richter haben, der vor dir sitzt und der eigentlich Rassist ist, oder der grad sich scheiden hat lassen und seine Frau hat ihn alles weggenommen oder umgekehrt eine Richterin und der Mann hat ihr alles weggenommen und sie ist jetzt männerfeindlich, egal wie.

Du kannst nicht in die Personen reinschauen, und das ist nie eine Entscheidung oder Meinung die sie offiziell sagen, aber natürlich fließt diese Wertung die du als Mensch hast in deine Entscheidungen mit ein, nur hast du andere Begründungen. Durch ein KI-System ist es aber so, dass viele dieser Vorurteile die wir haben, jetzt auf einmal sichtbar werden, weil nicht nur ein Einzelfall eine Entscheidung hat die vielleicht komisch ist, sondern sieht man dann auf einmal systematisch, dass bei Bewerbungen für Programmiererstellen Frauen benachteiligt werden. Sie könnte zeigen, dass schwarze Menschen in den USA bei kleineren Delikten härter bestraft werden als weiße Menschen bei schwereren Delikten. Oder eben bei der Kreditvergabe, dass Latinos die mehr verdienen als weiße, trotzdem den Kredit nicht kriegen. Wenn man jetzt auf einmal diese Ergebnisse hat, dann sehen wir das als Gesellschaft, und wir haben jetzt meiner Meinung nach die Möglichkeit, dass wir richtig agieren da dagegen. Wenn solche Ergebnisse einmal offenbart werden, dann können wir als Gesellschaft aktiv dagegen vorgehen. Es geht nicht darum, Einzelpersonen anzugreifen, sondern um das Erkennen von historischen Prägungen und systematischen Problemen. Die große

Chance besteht darin, dass wir Entscheidungen verbessern können, indem wir aufgrund von KI-Analysen auf Muster von Vorurteilen hinweisen.

Natürlich ist das kein einfacher Prozess. Menschen treffen oft subjektive Entscheidungen darüber, was die beste Wahl ist. Aber die Chance besteht darin, dass wir diese Themen angehen können, ohne Einzelpersonen anzugreifen. KI ermöglicht es uns, bewusst zu handeln und unsere Entscheidungsprozesse gezielt zu korrigieren. Das wird sicherlich nicht ohne Herausforderungen vonstattengehen, aber es bietet der Gesellschaft die Möglichkeit, positive Veränderungen anzustoßen. (…)

Eine weitere Chance besteht darin, hoffentlich viele der äußerst lästigen Aspekte des Lebens einfach beseitigen zu können. Momentan fühlt es sich so an, als würde alles immer mehr Arbeit bedeuten. Doch laut Goldman Sachs könnten durch KI angeblich 300 Millionen Arbeitsplätze wegfallen. Es wäre eine Erleichterung, all jene Aufgaben zu bewältigen, die einen eigentlich daran hindern, die Tätigkeiten zu verfolgen, die mit Leidenschaft ausgeführt werden - diejenigen, die das Kerngeschäft ausmachen. (…)

Grundsätzlich bin ich bereits überzeugt davon, dass KI in sämtlichen Bereichen, vor allem in Bezug auf Energieeffizienz und Optimierungspotenzial, enorme Potentiale bietet. Allerdings ist es bedauerlicherweise so, dass ich als Ethikerin und Juristin mich stets mit den negativen Aspekten auseinandersetze. Dennoch bin ich zutiefst davon überzeugt, dass wir einige dieser Herausforderungen bewältigen können. Hierzu zählt beispielsweise das Problem, dass ChatGPT täglich Kosten in Höhe von 700.000 € verursacht und unerschöpfliche Mengen an Energie und Ressourcen verbraucht. Doch ich bin zuversichtlich, dass wir diese Hindernisse überwinden können und ich bin eigentlich Tech-Optimistin.

Ich bin damit einverstanden,

☒ dass das Interview von Larissa Arthofer digital aufgezeichnet wird.

☒ dass das Interview von Larissa Arthofer transkribiert wird.

☒ dass ausgewählte Passagen des transkribierten Interviews in der VWA von Larissa Arthofer zitiert werden.

☒ dass das transkribierte Interview der VWA von Larissa Arthofer beigelegt wird.

☒ Mein Name darf in der VWA genannt werden

E-Mail: zehetmaiercarina@gmail.com

Vor- und Nachname : Mag. Carina Zehetmaier

Ort und Datum: Wien, 18.08.2023

Unterschrift: 

Als interviewende Person verpflichte ich mich zu einem ordnungsgemäßen Umgang mit den im Interview vorkommenden personenbezogenen Daten und dazu, die oben vereinbarten Rahmenbedingungen zu wahren.

Vor- und Nachname: Larissa Arthofer

Unterschrift: 

VWA INTERVIEW GESPRÄCHSPROTOKOLL

Women in AI Austria

Transkriptionskopf

Interviewkontext: Interview für die Erstellung meiner VWA,

Themenfeld „Korrelation zwischen Gehirnforschung und Künstlicher Intelligenz“

Name: Larissa Arthofer

Tag & Ort des Interviews: 24.08.2023, MS Teams Call

Dauer des Interviews: 60 Minuten

Name der interviewenden Person: Dr. DI. Mag.rer.soc.oec Isabella Hinterleitner M.Sc.

Im Fokus dieses Interviews steht Dr. DI. Mag.rer.soc.oec. Isabella Hinterleitner, M.Sc., deren umfassende Expertise in autonomen Fahrsystemen und Bilderkennung als Senior Researcherin bei Bosch sowie ihre Gründung von TechMeetsLegal sie zur idealen Expertin für meine VWA machen. Noch dazu vereint sie akademische Hintergründe aus Medizinischer Informatik, Wirtschaftsinformatik, Kognitionswissenschaften und einem Doktorat in Elektrotechnik, und ist zudem im Vorstand von Women in AI Austria aktiv.

Frage: Wie beeinflussen die Arbeitsweisen des menschlichen Gehirns die Funktionsprinzipien von KI-Algorithmen?

Nun, die Beziehung zwischen dem menschlichen Gehirn und KI-Algorithmen ist relativ gering. Ich habe mich sowohl mit Computer Science als auch mit Neurowissenschaften beschäftigt, und dabei ist mir klar geworden, dass wir es hier mit zwei sehr unterschiedlichen Dingen zu tun haben: dem menschlichen Gehirn, einem biologischen System, und KI, die auf komplett anderen Grundlagen basiert. Historisch gesehen gab es natürlich immer den Versuch, das Gehirn zu modellieren oder zu simulieren, was auch einer der Ansätze darstellt der zur Entwicklung von KI-Algorithmen geführt hat. Es geht darum, menschliche Kognition, die sehr einzigartig ist, zu verstehen, indem man sie durch eine andere Disziplin versucht abzubilden. Dafür nutzt man Informatik, also statistische und mathematische Verfahren, um verschiedene Phänomene wie Korrelationen, Regressionen oder Klassifizierungen zu analysieren. Die Frage, ob das eine das andere beeinflusst, ist meiner Meinung nach ziemlich gewagt, ähnlich der Sapir-Whorf-These in den Sprachwissenschaften, wo man sich fragt was zuerst kam: die Sprache oder das Denken. Ich glaube auch bei der KI ist es so, dass wenn du eine Beeinflussung findest, du ein Henne-Ei-Problem haben wirst, und dich immer fragen wirst was zuerst kam. Es ist ein bisschen wie das Henne-Ei-Problem.

Besonders interessant finde ich die Computational Neuroscience, ein Feld, das ich während meines Psycholinguistik-Studiums in England kennengelernt habe. Hier werden informatische Methoden eingesetzt, um neurowissenschaftliche Daten, wie PET-Scan-Ergebnisse, mittels Machine Learning zu

analysieren. Seit seiner Einführung um 2008 hat sich dieser Bereich stetig entwickelt. Trotz einiger Modelle, die von neuronalen Prozessen inspiriert sind, wie Perzeptronen-Modelle, der Hebb'sche-Lernregel und GANs (General Adversarial Network), ist die Übertragung menschlicher Kognition auf KI-Modelle begrenzt. Die Analogien zwischen Gehirn und KI enden oft bei diesen spezifischen Beispielen. Die Mehrheit der Machine-Learning-Techniken, bewegen sich weit entfernt von einer direkten Nachbildung menschlicher Gehirnprozesse.

Frage: Fallen dir andere nennenswerte Analogien zwischen Gehirnprozesssen und KI -Funktionen ein?

Reinforcement-Learning, genau, das fällt mir noch ein. Aber du siehst, es sind immer nur Anteile, die aus dem menschlichen Verhalten, aus unserer Natur, kopiert werden. Bei Reinforcement-Learning hast du die Bestrafungs- und Belohnungsfunktion, während beim Perzeptronenmodell die Aktivierungsfunktion im Fokus steht.

Zu etwas anderem: Ich habe im Bereich der Sprachwissenschaften auch ein Modell entwickelt, um die Verbreflexion bei Kindern zu untersuchen. Mich interessierte, wie sie irreguläre Verben, wie 'go, went, gone' konjugieren, speziell bei Kindern mit Williams-Syndrom, die bei der Bildung der irregulären Verben Schwierigkeiten zeigen und dazu neigen, ein 'ed' anzuhängen, statt die korrekte Form zu verwenden. Ausgehend dessen habe ich ein neuronales Netzwerk entwickelt, das genau diese Konjugationsmuster korrekt abbilden kann. Ich experimentierte damit, Teile des Netzwerks zu entfernen, um zu sehen, wie sich das auf die Ausgabe auswirkt, damit es 'go, goed, goed' anstatt 'go, went, gone' zu produzieren.

Dann habe ich den Ansatz umgekehrt und tatsächlich in der Praxis getestet, ich ging in verschiedene Krankenhäuser und arbeitete dort mit Kindern, nutzte Satzkästen und zeigte ihnen Bilder, um ihre Reaktionen zu beobachten. Diese Methode aus der Psychologie, obwohl mühsam, ermöglichte es mir, nach etwa einem halben Jahr die gewonnenen Erkenntnisse mit den Ergebnissen aus dem Modell zu vergleichen. Die nächste logische Frage wäre, ob ich mir vorstellen könnte, das in der Praxis anzuwenden. Genau da liegt der Kern: Wie entwickelt man neuronale Netzwerke, die die Verarbeitungsstrukturen des biologischen Gehirns nachahmen? Meine Forschung deutet darauf hin, dass besonders die Sprachverarbeitung ein relevantes Gebiet dafür sein könnte. Hierbei meine ich nicht die aktuellen Large Language Models, sondern die Grundlagen der Sprachproduktion und -konstruktion, die auf das Generieren einzelner Sätze und deren Kontextualisierung abzielen.

Das menschliche Gehirn und seine Komplexität bieten einen faszinierenden Vergleichspunkt zur Künstlichen Intelligenz. Mit meinem medizinischen Hintergrund erkenne ich, wie das Gehirn in unterschiedlichen Schichten funktioniert, insbesondere im Hinblick auf die Sprachzentren wie Broca und Wernicke. Broca ist für die Sprachproduktion zuständig, Wernicke für das Verständnis und die Semantik. Schädigungen in diesen Bereichen führen zu spezifischen Sprachstörungen: Ein Schlaganfall kann beispielsweise die Sprachproduktion beeinträchtigen, ohne das Sprachverständnis zu stören, sodass Patienten die Wörter kennen, sie aber nicht artikulieren können.

Im Vergleich dazu scheint KI weniger von solcher Symmetrie geprägt zu sein. KI-Systeme bauen auf der Addition einfacher Neuronen auf, um Komplexität zu erzeugen. Diese Ansätze zielen darauf ab, spezifische menschliche Fähigkeiten zu imitieren oder sogar zu übertreffen, aber sie sind weit davon entfernt, die Gesamtheit menschlicher Funktionen zu replizieren. Soft-KI kann einzelne Aufgaben, wie

Erkennungsaufgaben, besser als Menschen ausführen, aber Hard-KI, die menschliche Intelligenz in ihrer Ganzheit nachbildet, ist noch ein fernes Ziel. Wir können einzelne Teilfähigkeiten wie Sprachproduktion oder -verständnis modellieren, aber die Integration zahlreicher komplexer Funktionen in ein Netzwerk stellt eine enorme Herausforderung dar, die immense Rechenkapazitäten erfordert, wie sie vielleicht nur bei Einrichtungen wie CERN verfügbar sind.

Frage: Kannst du ein Beispiel aus der Bionik nennen, das zeigt, wie Technologien menschliche Gehirnfunktionen simulieren?

Also, in der Bionik gibt's eine Arbeit von einer Kollegin in der Elektrotechnik, die das menschliche Gehirn erklärt. Ihre Dissertation zeigt, wie Objekterkennung in technischen Systemen funktioniert. Die Autorin, hat 2006 promoviert. Ihre Forschung stellt die Verbindung zwischen Gehirnprozessen und Elektrotechnik her. Ich schick dir den Link zu ihrer Arbeit.
Was sie beschreibt, ist im Grunde der folgende Prozess. Du bekommst Input, ähnlich wie bei autonomen Fahrzeugen, die eine Menge Sensordaten verarbeiten müssen, um ein Objekt zu identifizieren. Diese Sensordatenfusion – also die Kombination von Bild, Akustik und Haptik – schafft ein bestimmtes Muster, das dann einer Klasse zugeordnet wird, zum Beispiel 'bewegte Objekte' oder noch spezifischer 'Fahrzeuge'.
Das funktioniert mit jedem Input, der reinkommt. Und in der bionischen oder technischen Umsetzung hast du dann für jede Funktion im Gehirn einen eigenen Baustein, ein eigenes Modul, das spezifisch dafür zuständig ist. Es sind nicht nur Schichten oder Layers, wie in manchen Modellen, sondern eher Module, die spezifische Aufgaben übernehmen, ohne dass dabei so viel auf Aktivierungsenergien gesetzt wird.

Frage: Wie funktioniert die Objekterkennung in autonomen Fahrzeugen, und welche Herausforderungen und Lösungen gibt es dabei?

Basierend auf meiner Arbeit im Bereich Innovation Management bei der Robert Bosch GmbH, speziell in Human-Computer Interaction für automotive Sensor-Systeme und in der kognitiven Robotik, kann ich bestätigen, dass die Objekterkennung – nehmen wir das Beispiel einer Kreuzung beim autonomen Fahren essenziell ist. Du hast verschiedene Inputdaten wie Bilddaten, Videodaten und möglicherweise Lidar-Daten (Light Detection and Ranging), die 3D-Informationen liefern. Das Zusammenlegen dieser Daten, besonders wenn sie redundant sind, verbessert die Sicherheit. Dann kommt der Schritt, aus diesen Daten Objekte zu extrahieren. Das ist eigentlich der kniffligste Teil, aber dafür gibt es bereits funktionierende Algorithmen, wie 'You only look once' (YOLO), das grid-basiert über das Bild läuft. Dieses Verfahren identifiziert zum Beispiel einen Zebrastreifen in einem komplexen Bild einer Kreuzung. Es muss sehr schnell gehen, in wenigen Millisekunden, damit das Fahrzeug aufgrund der erkannten Informationen rechtzeitig anhalten kann. YOLO markiert gefundene Objekte mit einem Rechteck und signalisiert so, was es erkannt hat. Manchmal muss diese Erkennung noch durch den Menschen bestätigt werden. Die Entwicklung geht natürlich stetig voran, ich glaube, wir sind schon bei Version 4, vor allem in Bezug auf die Schnelligkeit der Verarbeitung.

Frage: Wie schafft es unser Gehirn, Objekte so schnell zu erkennen und zu kategorisieren?

Das liegt daran, dass wir von Geburt an damit beginnen, Objekte zu erkennen und in Kategorien einzuteilen. Schon als Baby fangen wir an, die Welt um uns herum zu ordnen. Ich habe es selbst bei

meinem Kleinen gesehen – mit elf Monaten konnte mein Sohn schon einige Kategorien unterscheiden. Es ist faszinierend, denn obwohl ein zweijähriges Kind vielleicht nur 80 bis 100 Wörter sprechen kann, beherrscht es die Zuordnung von Objekten zu Kategorien viel früher. An der Uni Wien habe ich mit meinem Sohn an einer Studie teilgenommen, die untersucht, wie früh Kinder beginnen, Objekte zu kategorisieren, selbst bevor sie sprechen können.
Es ist eine enorme Herausforderung, diese Fähigkeiten bei kleinen Kindern zu testen, aber es ist entscheidend für unser Verständnis ihrer kognitiven Entwicklung. Deshalb ist es so wichtig, über längere Zeit Daten zu sammeln. Die frühe Bildung von Kategorien und die kontinuierliche Anwendung dieser Fähigkeiten im Laufe des Lebens führen dazu, dass wir immer besser werden. Von Anfang an erhalten wir diese Daten, und durch ständige Interaktion und Benennung von Objekten durch unsere Eltern und andere Bezugspersonen lernen wir.
Interessant ist auch, dass Kinder Sprache und Kategorien nicht einfach aus dem Fernsehen lernen können. Es muss eine echte Person sein, die mit ihnen interagiert, ohne Verzögerung. Das zeigt, wie wichtig direkte Kommunikation und Interaktion für die Sprachentwicklung und das Kategorienlernen sind. Die ersten sechs Monate sind entscheidend für die Erkennung der Muttersprache. Sogar der Dialekt wird erkannt, was zeigt, wie präzise unser Gehirn schon in den ersten Lebensmonaten arbeitet.

Frage: Glaubst du, dass eine KI mit ähnlich vielen Trainingsdaten wie Kinder in ihren ersten sechs Monaten gleich gute Ergebnisse in der Objekterkennung oder Sprachverarbeitung erzielen könnte?

Ja, wenn man einer KI ähnlich viele Trainingsdaten gibt, wie Kinder sie in den ersten sechs Monaten sammeln, stellt sich die Frage, ob sie genauso effektiv lernen könnte. Es gibt sicher schon Modelle, die das zu erreichen versuchen. Bei KI geht es darum, das richtige Modell auszuwählen. Ob du ein GAN nimmst oder was anderes, hängt davon ab, was du erreichen willst. Möchtest du nur das Sprachverständnis simulieren, oder zielen wir auf die viel komplexere Sprachproduktion ab? Die Sprachproduktion ist bei Menschen komplexer, da das Verständnis sich zuerst entwickelt, oft innerhalb des ersten Lebensjahres. Kinder sprechen manchmal bis zu ihrem zweieinhalbten Lebensjahr nicht, obwohl sie alles verstehen. In Bezug auf KI müssen wir also genau definieren, was wir simulieren oder trainieren wollen.
Übertrainierung, oder Overfitting, tritt auf, wenn ein KI-System zu genau auf die Trainingsdaten abgestimmt wird und die Fähigkeit verliert, auf neue Daten zu generalisieren. Wenn du deiner KI viele Daten gibst, besteht das Risiko, dass das Netzwerk übertrainiert wird und die Leistung dann stark nachlässt. Es gibt Mechanismen wie Dropout oder Overfitting-Schutzschichten, die ich in deinem CNN-Fallbeispiel auch gesehen habe, die verhindern sollen, dass das Netzwerk übertrainiert wird. Diese sind besonders wichtig, wenn wir versuchen, komplexe Fähigkeiten wie Bilderkennung zu trainieren.
Im menschlichen Gehirn gibt es jedoch nicht das Phänomen des Übertrainierens in dem Sinne, wie es bei KI-Systemen auftritt. Allerdings gibt es im Sport das Konzept des Burnouts, wenn zu intensiv trainiert wird, aber da rede ich wirklich von täglichem Training auf Wettkampfsniveau. Dies zeigt, dass es eine Grenze gibt, wie viel Training effektiv ist, bevor die Leistung wieder abnimmt. […]