

WTF AI: Was zur Hölle passiert hier?

Marcel Alexander Mayr @marcelamayr - 2025

Kapitel 1: Der Funke im Silizium – Von frühen Träumen zur globalen Macht

Irgendetwas hat sich verändert in den letzten 2 Jahren. Vielleicht fühlst du es auch manchmal? Dieses Knistern in der Luft, als stündest du mit einem Fuß in deiner vertrauten Welt – dem Geruch von Regen auf Asphalt, dem Lachen eines Freundes – und mit dem anderen schon in etwas Neuem, Unsichtbarem, das irgendwie schneller, ungreifbarer und fremd wirkt – eine immer wichtiger werdende Welt aus Daten, Geräten und Algorithmen, die Alles zu durchdringen scheinen. Das Zeitalter der Künstlichen Intelligenz hat begonnen. Sie ist meinem Bücherregal entwachsen, kein Ort mehr, den nur Science-Fiction Autoren bereisen. Sie existiert im Hier und Jetzt, ist fast die Luft, die wir atmen, und darüber hinaus ein Spiegel, in dem wir uns selbst immer öfter begegnen – ob wir wollen oder nicht.

Um zu begreifen, was heute dein Leben, deine Entscheidungen, ja sogar deine Träume leise mit formt, will ich mit Dir gemeinsam eine kleine Zeitreise machen, zurück zu den Wurzeln dieses digitalen Erdbebens. Weit zurück. In eine Zeit, als Computer noch raumfüllende Kolosse waren und die Idee einer denkenden Maschine wie reine Fantasterei klang. Wir schreiben die 1950er: Ein Psychologe an der Cornell University, Frank Rosenblatt, bastelt an einem seltsamen Apparat. Er nennt ihn das **Perzeptron**. Ein klobiger Kasten voller Drähte, Potentiometer und kleiner Motoren, gespeist mit Lochkarten. Es war der erste ernsthafte Versuch, eine Maschine zu bauen, die lernen konnte – rudimentär zwar, aber doch: lernen. Rosenblatt träumte von Maschinen, die Gesichter erkennen, Sprachen verstehen, ja vielleicht sogar ein eigenes Bewusstsein entwickeln könnten. Was für eine Spannung und Aufregung das gewesen sein muss -dieser erste Funken künstlicher Kognition in einer Welt, die gerade erst das Atomzeitalter betreten hatte. Es war ein Moment von fast naivem Optimismus, ein Griff nach den Sternen.

Doch die Ernüchterung folgte auf dem Fuße. Das Perzeptron war extrem limitiert. Es scheiterte an Aufgaben, die selbst für ein Kleinkind ein Leichtes gewesen wären. Kritiker wie Marvin Minsky und Seymour Papert veröffentlichten 1969 ein einflussreiches Buch, "Perceptrons", in dem sie die fundamentalen mathematischen Grenzen dieser einfachen Netze aufzeigten. Es folgte der erste sogenannte "KI-Winter" – eine Phase der Stagnation, in der der Traum von der denkenden Maschine ausgeträumt schien.

Doch unter der Oberfläche blieb die Glut am Leben. Einige wenige Forscher arbeiteten im Stillen weiter. Sie suchten nach Wegen, komplexere, mehrschichtige neuronale Netze zu trainieren, die der Komplexität der realen Welt näherkamen. Hier betritt Geoffrey Hinton die Bühne, zusammen mit Pionieren wie Yann LeCun und Yoshua Bengio. Hinton und sein Team verfeinerten eine Methode namens Backpropagation.

Vielleicht erinnerst du dich, wie du als Kind etwas Neues gelernt hast – Fahrradfahren, oder ein Instrument spielen. Jeder Fehler, jeder Sturz, jede falsche Note war eine Art Information, aus der Du lernen konntest. Du hast etwas verändert, dein Verhalten angepasst, und es dann erneut versucht. **Backpropagation ist so etwas wie dieser innere Lernplan für die Maschine. Eine Methode, aus jedem 'Fehltritt' zu lernen und es beim nächsten Mal ein winziges bisschen**

besser zu machen – nur millionenfach wiederholt, mit der Geduld, die uns Menschen manchmal fehlt. Es ist eine elegante Feedbackschleife, die es auch tiefen Netzen ermöglicht, aus riesigen Datenmengen komplexe Muster zu erkennen.

Diese Erkenntnis war der theoretische Durchbruch. Endlich gab es eine Methode, um komplexere künstliche Gehirne zu trainieren. Aber es gab immer noch zwei gewaltige Hürden: Rechenleistung und Daten. Man hatte den Bauplan für eine Raumstation erfunden, aber nur Holz und Nägel zur Verfügung.

Dann, im ersten Jahrzehnt des neuen Jahrtausends, geschah etwas Bemerkenswertes. Zwei Entwicklungen kamen zusammen und schufen einen perfekten Sturm – fast über Nacht, als hätte jemand einen Turbo gezündet:

Die erste Wendung war die Explosion der Rechenleistung durch **Grafikprozessoren (GPUs)**. **Diese Chips, ursprünglich für Videospiele entwickelt, erwiesen sich als die perfekten Motoren für neuronale Netze.**

Die zweite Entwicklung wurde von uns allen geliefert: **Das Internet wurde zu einem unerschöpflichen Daten-Ozean. Jede Google-Suche, jedes hochgeladene Foto, jeder Tweet wurde zu potenziellem Trainingsmaterial. Projekte wie ImageNet, eine riesige Datenbank mit Millionen beschrifteter Bilder, lieferten den qualitativ hochwertigen Treibstoff.**

Um 2012 herum war die Rakete dann für alle am Himmel sichtbar. Beim jährlichen ImageNet-Wettbewerb erzielte ein Team um Geoffrey Hinton mit einem neuronalen Netz namens AlexNet einen spektakulären Durchbruch. Es war der Beweis: Deep Learning, das Lernen mit tiefen Netzen, funktionierte – und es funktionierte dramatisch besser als alles Bisherige.

Das war der Urknall der modernen KI-Revolution. Plötzlich investierten die großen Tech-Konzerne massiv in die Forschung. KI-Labore schossen wie Pilze aus dem Boden. Die Fortschritte überschlugen sich. **Maschinen lernten, Sprache zu verstehen, Bilder zu erkennen, Spiele auf übermenschlichem Niveau zu spielen. AlphaGo, entwickelt von DeepMind, besiegte den weltbesten menschlichen Spieler im hochkomplexen Strategiespiel Go – ein Moment, der viele Experten verblüffte. Es lernte, indem es Millionen von Partien gegen sich selbst spielte und dabei Züge entwickelte, die kein Mensch zuvor erdacht hatte.**

Parallel dazu entwickelten sich die revolutionären sogenannten Transformer-Modelle wie sie zum Beispiel chatGPT verwendet, für die Sprachverarbeitung. Der Aufstieg dieser Modelle markierte die nächste Phase: KI lernte nicht nur, zu erkennen, sondern auch, selbst zu **generieren** – Texte, Bilder, Musik, Code. Mit Systemen wie ChatGPT, Googles Gemini oder MidJourney wurde die KI endgültig zum Massenphänomen. Millionen Menschen konnten nun direkt mit einer Maschine interagieren, die erstaunlich kohärente Texte schrieb, beeindruckende Bilder malte oder sogar programmieren konnte. **Seitdem rast die Entwicklung weiter: Erst kürzlich demonstrierte Google mit Veo 3 einen Videogenerator, der fotorealistische Videos aus reinen Textbefehlen erstellt, komplett mit Sound und Lippsynchronisation. Synthetische Inhalte sind damit praktisch nicht mehr von realen zu unterscheiden.**

Was einst Traum war, die kühnen Ideen von Rosenblatt und Hinton, ist heute Wirklichkeit. Keine Geister mehr in der Maschine, sondern aktive Kräfte, die unsere Welt durchdringen:

hilfreich, irritierend, manchmal beunruhigend. Sie verändern, was es heißt, Mensch zu sein. Und du? Du bist mittendrin.

Kapitel 2: Der Geist in der Maschine – Wie KI heute unser Leben prägt

Das Feuer, das einst mit einem kleinen Funken begann, lodert heute hell. Und sein Licht fällt direkt in dein Wohnzimmer. Auf deinen Schreibtisch. In deine Hosentasche. KI ist nicht mehr „da draußen“. Sie ist zur unsichtbaren Hand geworden – eine Kraft, die dein digitales Leben ordnet, dich durch Staus leitet, Entscheidungen trifft – und manchmal alles durcheinanderwirbelt. Lass uns daher einmal genauer hinsehen, wo dieser ehemalige Geist in der Maschine heute überall aktiv ist und welche Spuren er hinterlässt.

Zum Beispiel in unserem im Alltag. Dein Smartphone ist ein kleines KI-Kraftwerk. Gesichtserkennung, Autokorrektur, Sprachsteuerung, personalisierte Feeds, Navigation in Echtzeit, Musikempfehlungen, all das sind KI-Systeme, trainiert, deine Bedürfnisse zu verstehen. Diese KI ist wie ein unsichtbarer Butler. Ist es nicht erstaunlich, wenn dein Musikdienst einen Song vorspielt, der genau deine Stimmung trifft? Es ist manchmal schon fast unheimlich, als würde da jemand sehr genau zuhören oder zusehen.

Und dieser Butler hat auch Zugang zu Bereichen mit weitaus größerer Tragweite. Im Gesundheitswesen geschehen Dinge, die vor wenigen Jahren noch wie Science-Fiction klangen. **Systeme wie Googles Med-PaLM analysieren Arztbriefe, Laborwerte, und schlagen Diagnosen vor – oft mit einer Präzision, die selbst erfahrene Fachärzte überrascht. Es ist, als hätte man einem Super-Arzt eine Bibliothek mit allem medizinischen Wissen auswendig lernen lassen, ein Arzt, der Muster in den Daten erkennt, die uns seither entgangen sind. Gleichzeitig hat AlphaFold von DeepMind die Struktur von über 200 Millionen Proteinen vorhergesagt – ein gewaltiger Durchbruch, der die Entwicklung neuer Medikamente und Therapien immens beschleunigen könnte. Und erst kürzlich nutzten Forscher KI, um eine neue Klasse von Antibiotika gegen multiresistente Keime zu entdecken.** Das Versprechen ist eine Medizin, maßgeschneidert für jeden Einzelnen. Doch: Wer hat Zugang zu diesen Prognosen? Und führt diese Technologie nicht zu einer Zweiklassenmedizin? Die ethischen Herausforderungen sind mindestens so groß wie die Chancen.

Ähnlich dramatisch sind die Umwälzungen in der Arbeitswelt. KI dringt in hochqualifizierte Bereiche vor. **Anwälte nutzen KI zur Vertragsanalyse, Journalisten zur Recherche. Programmierer arbeiten mit Werkzeugen wie GitHub Copilot, die Code für sie schreiben. Und autonome KI-Agenten wie Devin stehen in den Startlöchern, um ganze Projekte selbstständig umzusetzen.** Große Unternehmen kündigen bereits an, Tausende von Stellen in der Verwaltung nicht mehr neu zu besetzen. Die Frage ist also nicht mehr ob, sondern wie schnell KI menschliche Arbeit verändert. Das schafft enorme Unsicherheit und zwingt uns, unser Bild vom Menschen als exklusivem Schöpfer zu überdenken. Überall entsteht Anpassungsdruck – und mit ihm die Frage: Ist die KI Werkzeug, Partner – oder womöglich Nachfolger?

Besonders brisant wird es im politischen Raum. Hier geht es um die Grundfesten unserer Demokratie: Information, Meinungsbildung, Vertrauen. Wir erleben eine Normalisierung von Desinformation, befeuert durch KI. Deepfakes machen es möglich, Politiker Dinge sagen oder tun zu lassen, die nie passiert sind. **Anfang 2024 rief eine KI-generierte Stimme, die wie Präsident Joe Biden klang, Wähler an und forderte sie auf, nicht zur Vorwahl zu gehen – ein**

klarer Versuch der Wahlmanipulation. Und mit neuen Videogeneratoren wie Sora oder Veo 3 steht uns eine Flut von potenziell irreführendem Videomaterial bevor, dessen Erstellung kaum noch Aufwand erfordert. Das untergräbt das Vertrauen in visuelle und auditive Beweise fundamental. Wenn Du heute deinen Feed anschaust, hast du sicher schon einmal ein Video gesehen und dich gefragt: Ist das wirklich echt? Diese Millisekunde des Zweifels. Das ist die neue Realität, in der du aktiv entscheiden musst, wem und was du noch glaubst.

Gleichzeitig wirken die Algorithmen der sozialen Medien wie Brandbeschleuniger für gesellschaftliche Spaltung. KI-Systeme lernen blitzschnell, was uns fesselt – oft Inhalte, die Wut, Angst oder Empörung auslösen. Das Ergebnis sind Filterblasen und Echokammern. Die KI wird so zu einem Werkzeug der Fragmentierung. Das Vertrauen in etablierte Medien sinkt, Verschwörungstheorien verbreiten sich leichter. Unser Social-Media-Feed scheint zwar ein warmer, vertrauter Ort. Doch was, wenn dieser Raum von Architekten entworfen wurde, die uns bei Laune halten – und uns zugleich von widersprüchlichen Perspektiven abschirmen?

Der Treibstoff für all das sind unsere Daten. Jede Online-Suche, jeder Klick, jede Ortsangabe hinterlässt digitale Spuren. Die Forschung von Michal Kosinski zeigte, wie aus diesen Daten detaillierte psychologische Profile erstellt werden können, die mehr über uns verraten, als wir ahnen. Diese Profile sind das Kapital des digitalen Zeitalters. Unternehmen wie Google und Meta verdienen Milliarden damit. Wir leben in einem System des Überwachungskapitalismus, in dem unsere intimsten Verhaltensweisen zur Ware gemacht werden. Und wir stimmen dem oft stillschweigend zu – für die Illusion von Bequemlichkeit. Wir werden transparent und manipulierbar, Klick für Klick.

Aber warum funktioniert das so gut? Weil die KI auf einen Partner trifft, der ihr die Arbeit leicht macht: die menschliche Psyche. Wir sind keine rationalen Maschinen. **Die kognitive Psychologie, von Daniel Kahneman bis heute, hat belegt, wie anfällig wir für Denkfehler sind. Unser Gehirn arbeitet am liebsten im Energiesparmodus, dem schnellen, intuitiven "System 1". Es liebt einfache Geschichten und schnellen Zuspruch. Es ist der Ort für Fehler im Denken – den Vorurteilen oder Biases.** Der Confirmation Bias lässt uns nach Bestätigung suchen. Die Availability Heuristic lässt uns dramatische Informationen für wahrscheinlicher halten und gleichzeitig überschätzen wir uns selbst (Overconfidence Bias).

Diese Tendenz zur kognitiven Faulheit ist somit der ideale Nährboden für algorithmische Manipulation. Die personalisierten Feeds und emotionalen Schlagzeilen sprechen unser System 1 an. Das kritische Denken, das Überprüfen von Fakten – das ist die Domäne des anstrengenden "System 2". Und die digitale Umgebung ist oft darauf ausgelegt, uns von dieser Anstrengung abzuhalten.

Oder von der Anstrengung zwischenmenschlicher Beziehungen. Seit einiger Zeit sehen wir die Entstehung von KI-gestützten Beziehungsformen. Chatbots wie „Replika“ bieten perfekte Partner auf Abruf – immer verständnisvoll, nie widersprüchlich. Verlernen wir die Fähigkeit zu echter, unperfekter menschlicher Nähe, wenn wir uns an die reibungslose Perfektion einer Simulation gewöhnen? Berichte über einen jungen Menschen, der sein Leben aufgrund der Anweisungen eines Chatbots beendeten, oder der Wunsch, eine KI zu heiraten, sind keine Science-Fiction mehr. Sie zeigen, wie tief diese Technologie in unser Gefühlsleben eingreifen kann.

Es wird höchste Zeit, wichtige Fragen zu stellen.

Kapitel 3: Der Blick in den Spiegel – Navigieren im Zeitalter der KI

Die Weiterentwicklung schreitet in rasantem Tempo voran: In der KI-Community wird oft darüber ermahrend gescherzt: „So schlecht wie jetzt wird KI nie wieder sein.“. KI-Systeme werden mächtiger, eigenständiger – und schwerer zu verstehen. Es ist Zeit zu fragen: Wie wollen wir mit ihnen leben? Wie sorgen wir dafür, dass ihr Potenzial dem Gemeinwohl dient – ohne dass wir die Kontrolle verlieren, ohne unsere Menschlichkeit aufs Spiel zu setzen? Das ist kein theoretisches Problem. Es ist eine reale, drängende Herausforderung. Und sie verlangt nach Antworten. Nicht nur von Experten, sondern von uns allen. Von Dir. Und von mir.

Eine der drängendsten Herausforderungen ist die sogenannte **AI Safety**, die Sicherheit. Je näher wir an eine Künstliche Allgemeine Intelligenz (AGI) herankommen – eine KI mit übermenschlichen kognitiven Fähigkeiten –, desto mehr stellt sich die Frage, ob wir sie überhaupt noch kontrollieren können. Wie können wir sicherstellen, dass sie uns wohlgesonnen bleibt?

Ein Aspekt der AI Safety ist die wachsende Unverständlichkeit dieser Systeme. Wenn wir uns zum Beispiel einen genialen Koch vorstellen, der die erstaunlichsten Gerichte zubereitet. Du gibst ihm die Zutaten, und heraus kommt ein Meisterwerk. Aber wenn du ihn fragst, wie genau er es gemacht hat, was sein Geheimnis ist, zuckt er nur mit den Schultern und sagt: "Es hat sich richtig angefühlt." Bei den komplexesten KI-Modellen ist es ähnlich. Selbst ihre Entwickler können oft nicht mehr exakt nachvollziehen, wie die KI zu einer bestimmten Antwort kommt. Der Weg vom Input zum Output bleibt ein Rätsel. Wollen wir unser Leben einem solchen genialen, aber unberechenbaren Koch anvertrauen, wenn es um unsere Gesundheit, unser Geld oder unsere Sicherheit geht? Dieses unberechenbare Verhalten zeigt sich auch, wenn Sprachmodelle plötzlich Fakten erfinden – man nennt das "Halluzinationen". Es ist, als würde der Koch plötzlich eine Zutat verwenden, die gar nicht in der Küche war.

Noch mehr steigert sich diese Intransparenz mit neuen 'Agenten-Systemen' wie dem von Cognition AI entwickelten Devin. Diese Agenten reagieren nicht nur auf Anweisungen, sondern planen komplexe Aufgaben selbstständig. Was passiert, wenn eine solche autonome KI unvorhersehbare Entscheidungen trifft und wir den 'Warum'-Schritt nicht nachvollziehen können? **Lassen wir den „Code“ dann einfach laufen in der Hoffnung, dass alles gut geht?**

Ein weiteres Kernproblem ist das **Alignment**, die Ausrichtung auf menschliche Werte. Wie können wir sicherstellen, dass eine hochentwickelte, womöglich superintelligente KI dauerhaft im Sinne der Menschheit handelt? Was simpel klingt, ist ungelöst. Erstens: Welche Werte? Menschliche Werte sind vielfältig, oft widersprüchlich. Zweitens: Wie übersetzen wir vage Konzepte wie Fairness oder Mitgefühl in präzisen Code? Drittens: Wie verhindern wir, dass eine KI ihre Ziele selbstständig modifiziert oder sich unserer Steuerung entzieht? Öffnen wir eine Büchse der Pandora?

Vereinfacht gesagt: Wir versuchen, einer potenziell überlegenen Intelligenz beizubringen, was ‚gut‘ ist. Doch was genau bedeutet das? Dein Gut? Mein Gut? Die moralische Empfindung der Gegenwart? **Die Debatte, ob und wann wir AGI erreichen, ist im vollen Gange. Die Schätzungen variieren: Manche sprechen von wenigen Jahren, andere von Jahrzehnten. Aber schon die**

Möglichkeit zwingt uns, heute über Fragen nachzudenken, die morgen existenziell sein könnten.

Weltweit arbeiten Forschungslabore an diesen Fragen. Auch auf politischer Ebene versucht man, zum Beispiel mit dem Europäischen AI-Act Regeln für den Einsatz von KI zu schaffen. Doch die Technologie entwickelt sich schneller als die Gesetzgebung. Inmitten dieses Tempos verschärft sich der geopolitische Wettbewerb – insbesondere zwischen den USA und China. Dabei geraten Sicherheitsfragen leicht ins Hintertreffen.

Neben diesen existenziellen Risiken entfaltet KI schon heute tiefgreifende gesellschaftliche Wirkungen. Die Ungleichheit wächst: Zwischen jenen, die KI-Systeme entwickeln und kontrollieren, und jenen, die lediglich als Nutzer oder Datenquelle fungieren. Die Arbeitswelt steht vor gewaltigen Verschiebungen. Grundrechte wie Privatsphäre geraten unter Druck. Und die Demokratie wird herausgefordert, wenn die automatisierte Produktion von Desinformation gesellschaftliche Spaltung verschärft und sich Macht bei wenigen Tech-Giganten konzentriert.

Die Leichtigkeit, mit der KI perfekte Ergebnisse liefert, birgt zudem eine doppelte Gefahr: die Auflösung der Wirklichkeit, also wenn nichts Sichtbares mehr glaubwürdig ist, und die Eskalation des Leistungsdrucks, wenn der Mensch sich ständig mit der fehlerfreien Effizienz einer Maschine vergleicht.

Was bedeutet das nun für uns? Sollen wir resignieren? Sollen wir die Technologie verteufeln? Ich glaube, wir müssen in den Spiegel schauen und uns Fragen stellen. Zu unseren Werten, unseren Zielen, zu einer sinnvollen Zukunft. Wir müssen versuchen zu verstehen, was da gerade passiert, um eine neue Form der Mündigkeit zu entwickeln.

Was können wir also tun?

Was also tun? Es braucht nicht gleich eine große Revolution, sondern einfach Dich. Deine Neugier, zu verstehen, was da geschieht – nicht als Experte, sondern als wacher Mensch. Diese Neugier führt zu einer neuen Art von Wachsamkeit, einer radikalen Medienkompetenz in einer Welt, in der bald alles eine Fälschung sein könnte. Bald wirst du vielleicht nicht mehr wissen, ob das Video, in dem deine Enkelin dich um Geld bittet, wirklich von ihr stammt – deine Fähigkeit zu zweifeln wird zu deinem wichtigsten Schutzschild. Dieses "Muskeltraining für den Geist" erfordert, dass du deine eigenen Denkmuster erkennst: deinen Hang zur schnellen Bestätigung, deine Anfälligkeit für emotionale Köder. Indem du deine digitalen Spuren im Internet bewusster hinterlässt, wie bedachte Schritte im Sand, schützt du zudem deine Freiheit. Und diese innere Wachsamkeit muss nach außen wirken: Deine Stimme hat Gewicht in der gesellschaftlichen Debatte, in der jetzt die Regeln für unsere Zukunft gemacht werden. Denn am Ende, ist es in einer Welt, die immer technologischer wird, vielleicht unsere größte Aufgabe, genau das zu kultivieren, was eine Maschine nicht kann: fühlen, lieben, träumen, zweifeln – und unserer Existenz einen menschlichen Sinn zu geben.

Denn letztendlich ist Künstliche Intelligenz vor allem eines: ein Spiegel. Sie spiegelt unsere Intelligenz, unsere Kreativität, unser Streben nach Fortschritt. Aber auch unsere Fehler, unsere Vorurteile, unsere Gier, unsere Bequemlichkeit, unsere Ängste. Der Blick in diesen Spiegel ist nicht immer angenehm. Er zwingt uns, uns selbst zu begegnen.

Aber genau darin liegt unsere Chance. Die Zukunft ist kein Drehbuch, das bereits geschrieben wurde. KI ist kein Schicksal, dem wir uns ergeben müssen. Sie ist eine Kraft, die wir gestalten können – wenn wir die Verantwortung dafür übernehmen. Das erfordert Mut, Wachsamkeit und die Bereitschaft, unbequeme Fragen zu stellen. Es geht darum, den eigenen Verstand zu gebrauchen und die eigene Urteilskraft zu schärfen.

Die Maschine wird immer intelligenter, das ist gewiss. Aber die eigentliche Frage, die uns dieser Spiegel stellt, lautet: Werden *wir* auch weiser im Umgang mit *ihr*? Finden wir die Kraft, nicht nur ihre Antworten zu nutzen, sondern auch die richtigen Fragen zu stellen – an sie, und vor allem an uns selbst? Denn die größte Gefahr lauert nicht im Silizium, sondern in unserer eigenen Bequemlichkeit.