

Positionspapier der EvoTegra GmbH zur aktuellen Diskussion um Ethik und Erklärbarkeit in Bezug auf künstliche Intelligenz (KI)

Die EvoTegra GmbH ist Anbieter von kundenspezifischen Lösungen der Bildverarbeitung auf Basis von künstlicher Intelligenz und Deep Learning. Mit unseren individuellen Lösungen sprechen wir gezielt kleine und mittlere Unternehmen an, die überwiegend nicht die Möglichkeit haben eigenes KI Knowhow aufzubauen.

Ethik ist eine „philosophische Disziplin oder einzelne Lehre, die das sittliche Verhalten *des Menschen* zum Gegenstand hat.“. Durch Ethik wird eine kollektive subjektive Bewertung zu einer sozialen Norm, nicht jedoch zu einer objektivierbaren Größe. Dies lässt sich unter anderem daran erkennen, dass ethische Bewertungen teilweise von einem Kulturkreis abhängig sind und grundsätzlich einem zeitlichen Wandel unterliegen.

Die heute in der Praxis verwendete „schwache künstliche Intelligenz“ ist eine sich selbst optimierende mathematische Annäherung an eine unbekannte komplexe Funktion. Die Optimierung erfolgt anhand von Beobachtungen, die der künstlichen Intelligenz in Form von Daten während der Trainingsphase zur Verfügung gestellt werden. Während der anschließenden Anwendungsphase lernt eine schwache KI nicht mehr dazu.

Da heutige künstliche Intelligenz zu keinem Zeitpunkt ein Bewusstsein über das Erkannte entwickelt, ist KI dabei prinzipiell vollkommen objektiv. Die Risiken aus der Anwendung von KI liegen daher in den verwendeten Daten und nicht in der KI selbst.

Mit der aktuellen Diskussion von KI und Ethik soll verhindert werden, dass Systeme der künstlichen Intelligenz Menschen diskriminieren. Dazu sollen diese Systeme so entwickelt werden, dass man die Gefahr unbeabsichtigter Folgen minimiert. Ein Unternehmen müsste vor dem Einsatz dann beispielsweise testen, ob die künstliche Intelligenz bestimmte Personengruppen benachteiligt oder abwertet. ([Link](#))

In welchem Zusammenhang kann Ethik und KI gesehen werden?

1.) Integration von Ethik in KI

Da KI kein Bewusstsein über den bewerteten Inhalt entwickelt und Ethik nicht objektivierbar ist, lässt sich Ethik nicht sinnvoll in künstliche Intelligenz integrieren.

2.) Ethik der Daten

Das Ergebnis einer KI basiert auf den Daten. Daher ist ein Ansatz für den ethischen Umgang mit KI der verantwortungsvolle Umgang mit den Daten.

3.) Ethik der Ergebnisse

Hier stellt sich die Frage, wie das Ergebnis einer KI zu bewerten ist und wie KI und Mensch so interagieren, dass die ethischen Grundsätze gewahrt werden.

Vergleicht man künstliche Intelligenz mit den aktuellen Methoden stellt man fest, dass KI prinzipiell den gleichen Zweck erfüllt wie die wie „Elektronische Datenverarbeitung“ seit rund 60 Jahren. Daten werden verarbeitet und bewertet. Die beiden Ansätze unterscheiden sich lediglich in der Art der Modelle, die zu Bewertung der Daten verwendet werden.

Der Mensch ist in seiner Fähigkeit komplexe Zusammenhänge zu erkennen, durch die Millersche Zahl (7 ± 2) begrenzt. In menschlichen Modellen werden daher komplexe Zusammenhänge vereinfacht und anschließend z.B. in Form eines Entscheidungsbaumes bewertet.

Daraus ergibt sich eine Erklärbarkeit innerhalb des Modells. Die Frage, ob die Annahmen des Modells statistisch valide sind oder (implizit) diskriminiert wird jedoch in der Regel weder gestellt noch geprüft. Es ist weitgehend gesellschaftlich akzeptiert eine *Begründung* zu bekommen, deren *Erklärbarkeit* d.h. Relevanz in Bezug auf die Entscheidung in der Regel jedoch nicht überprüfbar ist. Dabei zeigt sich gerade durch den Einsatz von KI, dass menschliche Annahmen oft unvollständig oder teilweise falsch sind.

Modelle der künstlichen Intelligenz hingegen erkennen Zusammenhänge direkt in komplexen Daten. Statt vereinfachter Annahmen spiegeln sie daher die statistische Realität der Daten. Da die Bewertung grundsätzlich anhand statistischer Zusammenhänge erfolgt, ermöglicht dies eine prinzipiell gerechtere Bewertung. Problematisch wird dies jedoch, wenn Daten unvollständig sind oder scheinbare Zusammenhänge enthalten, die nicht auf die Realität übertragbar sind.

Aufgrund der rechtlichen Risiken ist es ein grundsätzliches Ziel Systeme so zu entwickeln, dass Menschen nicht diskriminiert werden.

Um dies zu verhindern, gibt es im Umgang mit KI bereits heute zahlreiche Ansätze wie z.B.:

1. Entfernung von potentiell diskriminierenden Merkmalen aus den Daten
2. Ergänzung von unterrepräsentierten Datenmerkmalen
3. Höhere Gewichtung von unterrepräsentierten Datenmerkmalen

Insgesamt handelt es sich um ein alltägliches und in der Regel gut beherrschbares Problem. KI bietet jedoch keine universelle Lösung für jede Aufgabenstellung. Sind grundsätzlich keine ausreichenden Daten vorhanden, stellt KI in der Regel keinen sinnvollen Lösungsansatz dar.

Menschliche Intuition und Ethik ergänzt sich perfekt mit der Geschwindigkeit und Objektivität der KI. Da KI falsche oder unvollständige Annahmen erkennt, lernt im Zusammenspiel nicht nur die KI vom Mensch, sondern der Mensch auch von der KI. Da menschliche Intuition auch mittelfristig nicht durch künstliche Intelligenz ersetzt werden kann, empfehlen wir grundsätzlich den Einsatz von KI als Assistenzsystem. Damit lässt sich nicht nur die Qualität und die Effizienz eines Unternehmens steigern. Unternehmen können hier in Zeiten des Fachkräftemangels besonders profitieren indem KI verwendet wird um die Produktivität von Schlüsselmitarbeitern zu steigern.

Rechtliche Risiken durch fehlende bundesweite Richtlinien im Umgang mit der DSGVO stellen bereits heute ein erhebliches Hindernis bei der Einführung von künstlicher Intelligenz dar. Die Verwendung von KI in einem *produktiven* Anwendungsfall innerhalb der Breite der Industrie oder dem öffentlichen Sektor ist aus unserer Sicht immer noch ein Einzelfall. Die Forderung das Systeme der künstlichen Intelligenz *grundsätzlich* ethischen Grundsätzen entsprechen und erklärbar sein müssen, bedeutet daher eine zusätzliche Hürde, die auch von der heutigen klassischen Datenverarbeitung in der Regel nicht erfüllt wird.

Unabhängig von der Technologie dürfen Menschen nach Art. 22 DSGVO nicht vollautomatischen Entscheidungen unterworfen werden, sofern die Entscheidung erhebliche Auswirkungen auf die Person hat. Da so die Einhaltung ethischer Grundsätze und die Erklärbarkeit bei relevanten Entscheidungen grundsätzlich gegeben sein muss, empfehlen wir in der Betrachtung eine Gleichstellung von Systemen der künstlichen Intelligenz und der klassischen Datenverarbeitung.

Um den enormen Rückstand Deutschlands bei der künstlichen Intelligenz nicht weiter zu vergrößern sollte aus unserer Sicht das Ziel der weiteren Diskussion sein, bundesweit verbindliche rechtliche Rahmenbedingungen für den Einsatz von KI zu schaffen, Bürokratie in Bezug auf die Förderung zu verringern und ausreichende Mittel bereit zu stellen sowie die Aufklärung innerhalb der Industrie und des öffentlichen Sektors über die *konkreten* Einsatzmöglichkeiten von KI zu verbessern.