# xAI / Explainable AI: Schlüssel zur Transparenz komplexer KI-Anwendungen – die KI-Serie (Teil 1)

**Die Bedeutung von künstlicher Intelligenz (KI/AI) wächst stetig, auch und gerade in der Finanzbranche. Die dabei verwendeten Prognosemodelle liefern immer genauere Resultate, sie werden dabei aber auch immer komplexer. Wenn dadurch das von der KI-Anwendung berechnete Ergebnis für den Nutzer oder den Betroffenen nicht mehr nachvollziehbar ist, leidet die Akzeptanz der Technologie insgesamt. Hier hilft nur eins: Erklären! Aber wie? xAI!**

***von Marc-Nicolas Glöckner, Senior Manager und Jan Eßer, Senior Consultant PPI***

Marc-Nicolas Glöckner, Senior Manager PPI Quelle: PPI

Künstliche Intelligenz gilt als eine der großen

Jan Eßer, Senior Consultant PPI Quelle: PPI

Zukunftstechnologien des digitalen Zeitalters. Von der Anwendung versprechen sich die Unternehmen schnellere und bessere Problemlösungen, beschleunigte Prozesse und geringeren Ressourcenverbrauch. Hochkomplexe KI-Modelle sind zwar inzwischen in der Lage, Antworten auf komplizierte Fragestellungen zu geben. Aber diejenigen, die mit den Rückmeldungen der Maschine arbeiten müssen, fragen sich nicht selten, wie das Ergebnis zustande kam. Bei einer Umfrage des Branchenverbandes Bitkom aus dem Sommer 2022 sahen die Hälfte der Unternehmen die mangelnde Nachvollziehbarkeit der Resultate als Risiko beim Einsatz von KI (https://www.bitkom.org/Presse/Presseinformation/Kuenstliche-Intelligenz-2022).

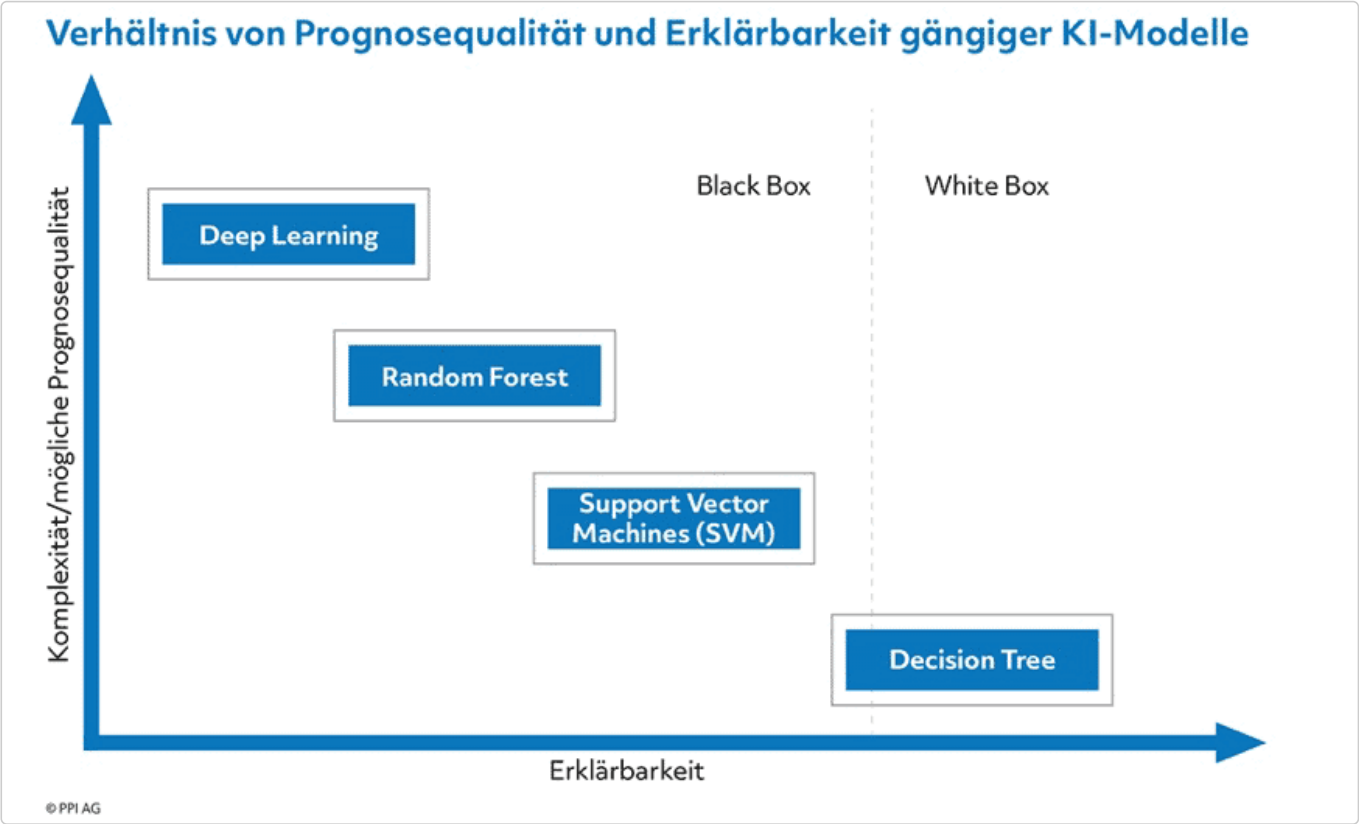
## Transparenz ist erfolgsentscheidend

Wenn fast die Hälfte der Marktteilnehmer in Deutschland bei KI-Algorithmen die fehlende Transparenz als Risiko sehen, sollten Entwickler und Förderer der Technologie dies als Alarmzeichen verstehen. Denn wenn die Entscheidung der KI als solche nicht erklärbar ist, schwindet die Akzeptanz der Ergebnisse, die Skepsis

gegenüber der Technologie insgesamt steigt. Gleichzeitig ist der Einsatz der Technologie für das Bestehen im internationalen Wettbewerb unabdingbar. Deshalb ist es wichtig, die Resultate von KI-Anwendungen transparent und verständlich erklären zu können.

## Problem Black Box

Dies ist aber einfacher gesagt als getan. Unkomplizierte, lineare Algorithmen wie etwa klassische Decision Trees sind für die meisten Menschen nachvollziehbar. Allerdings eignen sich darauf basierende Anwendungen in der Regel auch nur für simple Fragestellungen. Einfachere Chatbots mögen damit noch auskommen – sobald die Zahl der zu berücksichtigenden Faktoren steigt und komplexere Modelle gefragt sind, wird es mit der Nachvollziehbarkeit mühsam. Gerade sehr wirkmächtige Algorithmen gelten als Black Box.




Quelle: PPI

## Zu komplex für einfache Erklärungen

Prinzipiell lassen sich alle Rechenschritte von Algorithmen nachvollziehen und in Formeln darstellen. Aber die oft sehr hohe Anzahl von Zwischenschritten in den Berechnungen und die fehlende Linearität auf dem Weg zu einer Entscheidung führen zu einer Komplexität, die im Bereich nichtlinearer Deep-Learning-Anwendungen astronomische Ausmaße erreichen kann. Einzelne Entscheidungen durch Darstellungen der verwendeten Formeln aufzuschlüsseln, ist schlicht nicht möglich. Trifft eine Anwendung aus dieser Black Box heraus eine Entscheidung, dann stellt sich daher dem Nutzer immer die Frage: „Wie kommt die Maschine jetzt ausgerechnet zu diesem Ergebnis?“ Dies insbesondere dann, wenn die Antwort auf die Ausgangsfrage anders ausfällt als erwartet oder aber der Mensch gar nicht zur Beantwortung in der Lage ist.

## Explainable AI wird zur Wissenschaft

Dieses Dilemma aufzulösen, ist eine Aufgabe, mit der sich inzwischen ein ganzes Forschungsfeld befasst, das als explainable AI oder kurz xAI bezeichnet wird. Mit der Zunahme neu entwickelter KI-Modelle in den vergangenen Jahren stieg auch die Anzahl der Veröffentlichungen, die das Verhalten der Algorithmen erklären sollen. Spätestens seit 2016 ist ein sprunghafter Anstieg bei den wissenschaftlichen Publikationen zu verzeichnen.

SERIE: explainable AI (xAI)

**Teil 1:** **Mitentscheidend für den Erfolg von KI-Anwendungen: explainable AI**

**Teil 2:** Vorhersagen eines KI-Modells einfach akzeptieren?

**Teil 3:** Methoden und Techniken der xAI im Detail

Dabei nützt xAI zunächst einmal den Entwicklern der KI-Modelle. Denn sie können durch die Anwendung von xAI-Ansätzen feststellen, worauf etwaige Fehleinschätzungen basieren. Beispielsweise können dadurch Algorithmen angepasst oder Modelle neu trainiert werden. Dann kann es unter Umständen notwendig werden, den Algorithmus für den Einsatzbereich neu zu trainieren.

## Begründung mit Mehrwert

Aber auch der Nutzen einzelner Prognosemodelle erhöht sich durch xAI. Schließlich lassen sich deutlich leichter Maßnahmen für die Zukunft ableiten, wenn die Entscheidungsgründe einer KI-Anwendung bekannt sind. So ist die beispielhafte Aussage „der Umsatz mit dem Kreditprodukt Familien-Komfort-Kredit wird im nächsten Quartal um 15 Prozent sinken“ für sich genommen nur die Beschreibung einer prognostizierten Entwicklung. Lautet das Ergebnis aber „der Umsatz des Kreditprodukts wird um 15 Prozent sinken, da die Sichtbarkeit des Angebots auf Vergleichsplattformen sinkt“, können die Verantwortlichen die mit Hilfe von xAI identifizierte Hauptursache klar benennen.

Entsprechend anspruchsvoll sind die Erklärungsansätze, insbesondere für Black-Box-Modelle. Abhängig von dem Algorithmus, dem betroffenen Einsatzfeld und anderen Faktoren gibt es eine Vielzahl denkbarer Vorgehensweisen, um die nötige Transparenz zu schaffen. Diese Methoden werden im letzten Teil der vorliegenden Serie vertieft behandelt.

## Nachvollziehbarkeit ist auch rechtlich relevant

Darüber hinaus ist xAI auch für Juristen interessant. Denn diese sind absehbar gefordert, wenn Finanzdienstleister aufgrund einer KI nachteilige Entscheidungen für Kunden treffen. Dann muss im Zweifel vor Gericht bewiesen werden, dass es – beispielsweise für die Ablehnung eines Kredites – nachvollziehbare Gründe gab. Die bloße Aussage „das hat unsere Software gesagt“ dürfte keineswegs ausreichen. Auch die Aufsicht hat diese künftige Problematik bereits erkannt und beginnt, den Einsatz von KI zu regulieren. So hat die Bundesanstalt für Finanzdienstleistungsaufsicht, ausgehend von Initiativen der Vereinten Nationen und der EU, mittlerweile ein Prinzipienpapier zu diesem Thema herausgegeben. Viele der dort aufgeführten Forderungen sind ohne xAI kaum erfüllbar. Wie der Gesetzgeber das Thema insgesamt sieht und wo die Grenzen des KI-Einsatzes liegen, behandelt der nächste Teil der Serie zum Thema explainable AI.

Marc-Nicolas Glöckner PPI und Jan Eßer PPI ∎

**Im zweiten Teil (Montag) unserer Serie ‘Explainable AI (xAI)’ beleuchten wir, warum ein tiefgreifendes Verständnis von Algorithmen entscheidend ist und welche regulatorischen Aspekte dabei berücksichtigt werden müssen.**