

Ein eindrückliches Beispiel für die **Blackbox-Kapitulation** stammt aus der medizinischen KI-Diagnostik, genauer gesagt aus der Intensivmedizin. Hier kollidiert das menschliche Bedürfnis nach Logik direkt mit der unbegreiflichen Komplexität moderner Deep-Learning-Systeme.

Das Szenario: Der prädiktive Alarm auf der Intensivstation

Stell dir eine Intensivstation vor. Ein Patient liegt im Bett, überwacht von Dutzenden Sensoren. Ein hochmodernes KI-System analysiert kontinuierlich und in Echtzeit Tausende Datenpunkte: Herzfrequenzvariabilität, Blutdruckkurven, Sauerstoffsättigung, Atemfrequenz und Laborwerte der letzten Tage.

Das System wurde mit den Daten von Millionen Patientengeschichten trainiert. Seine Aufgabe: Erkennen, ob ein Patient in den nächsten 24 Stunden einen plötzlichen, lebensbedrohlichen Herzstillstand erleiden wird.

1. Das Ereignis

Plötzlich schlägt die KI Alarm. Sie meldet: *„Patient Müller wird mit einer Wahrscheinlichkeit von 94 % in den nächsten 12 Stunden einen Herzstillstand erleiden.“*

2. Der menschliche Kontrollversuch

Die Ärzte und Pfleger eilten zum Bett von Patient Müller. Sie untersuchen ihn, schauen auf die Monitore. Das Ergebnis: **Der Patient wirkt stabil**. Die klassischen medizinischen Warnzeichen (akuter Blutdruckabfall, Herzrhythmusstörungen) fehlen im Moment.

Der Chefarzt fragt nun die KI (bzw. sucht in der Software nach der Erklärung): *„Warum? Welche Variable hat den Ausschlag gegeben? Ist es der Kaliumwert? Ist es die Atemfrequenz?“*

3. Die Blackbox-Realität

Die Antwort der Software ist – technisch bedingt – eine mathematische Leere. Die KI hat keine einfache medizinische Regel angewandt wie: *„Wenn Wert A steigt und Wert B fällt, dann passiert C.“* Stattdessen hat sie die Daten durch ein neuronales Netzwerk mit hunderten Millionen von Parametern und Schichten gejagt. Sie hat minimale, für das menschliche Auge unsichtbare Korrelationen in den Millisekunden-Abständen des Herzschlags in Kombination mit einem minimalen Temperaturabfall von vor drei Stunden und der Frequenz der Medikamentengabe erkannt.

Das System kann dem Arzt kein medizinisches „Warum“ nennen. Es kann nur sagen: *„Die Muster in den Daten ähneln zu 94 % den Mustern von Patienten, die kurz darauf gestorben sind.“*

Die Kapitulation des Experten

Die Ärzte stehen nun vor einem ethischen und logischen Dilemma. Sie haben zwei Optionen:

1. **Die KI ignorieren**, weil sie die Entscheidung medizinisch nicht nachvollziehen können. Stirbt der Patient dann, haben sie ein Menschenleben auf dem Gewissen, obwohl sie gewarnt wurden.
2. **Der KI blind vertrauen** und vorbeugende, potenziell riskante und invasive Behandlungen einleiten, ohne zu wissen, ob sie überhaupt notwendig sind oder das Problem sogar verschlimmern.

Da die klinischen Studien gezeigt haben, dass diese KI in 95 % der Fälle recht hat (weit treffsicherer als jeder menschliche Arzt), **kapituliert der Chefarzt**. Er sagt: *„Ich verstehe nicht, was die Maschine sieht. Es ergibt medizinisch für mich gerade keinen Sinn. Aber wir behandeln den Patienten jetzt trotzdem auf Verdacht eines Herzstillstands.“*

Warum das ein Problem ist: Hier findet eine schleichende Entmachtung des menschlichen Sachverstands statt. Der Mensch kapituliert vor der Komplexität der Maschine und mutiert vom „Entscheider“ zum bloßen „Ausführenden“ eines Algorithmus, den niemand mehr versteht. Wenn die KI in diesem Fall falsch liegt (weil sie vielleicht nur ein Störsignal im Stromnetz der Klinik als Muster fehlinterpretiert hat), merkt es niemand mehr.