



Explainable Artificial Intelligence (XAI)

# Wenn sich die KI selbst erklären muss

Wissenschaftler der Mayo Clinic in Florida haben ein KI-System entwickelt, das innerhalb von zehn Sekunden etwaige Herzrhythmusstörungen mit wiederkehrendem Vorhofflimmern selbst dann erkennen kann, wenn die EKG-Signale während der Aufnahme für den Arzt noch unauffällig sind.

**E**s kann auf einem Smartphone zum Monitoring von Risikogruppen für Schlaganfall, Bluthochdruck oder Diabetes laufen, bedarf aber noch weiterer klinischer Prüfungen.

Der Algorithmus wurde in einem neuronalen Netzwerk mittels Deep Learning an 650.000 EKG-Aufnahmen trainiert. Er erkennt die Störungen anhand subtiler Veränderungen in der P-Welle mit einer Sensitivität von 79 Prozent und einer Spezifität von 79,5 Prozent, die sich durch mehrere EKG-Aufnahmen noch steigern lassen. Wie der Algorithmus seine Erkenntnis gewinnt, können die Medizinforscher aber selbst nicht sagen.

Diese Unklarheit stellt aber Entwickler, Ärzte und auch Patienten vor drei Probleme. Zunächst muss der Entwickler seine eigene Voreingenommenheit während der Lernphase möglichst ausschalten. So wurden KI-Algorithmen in der Dermatologie v.a. mit Hautkrebsbildern von hellhäutigen Patienten trainiert. Sie versagten häufig bei dunkelhäutigen Menschen. Für den Arzt selbst ist der Weg von den Patientendaten zu den KI-Diagnosen völlig unklar. Wie will er dann Fehler im Entscheidungsprozess aufdecken? Drittens könnte diese Unkenntnis das Vertrauensverhältnis des Patienten zum Arzt stören, wenn er blind einem Urteil des Computers vertrauen soll. Der Einsatz von intransparenten KI-Blackboxen provoziert zudem noch

Konflikte mit der Datenschutz-Grundverordnung (DSGVO). Artikel 22 formuliert das Recht des Patienten, „nicht einer ausschließlich auf einer automatisierten Verarbeitung – einschließlich Profiling – beruhenden Entscheidung unterworfen zu werden“, die gegenüber Betroffenen „rechtliche Wirkung entfaltet oder sie in ähnlicher Weise erheblich beeinträchtigt“. Auch haben Betroffene wegen des Transparenz-Grundsatzes in Artikel 15.1(h) das Recht, „aussagekräftige Informationen über die involvierte Logik“ dieser Entscheidung zu erhalten.

Die Beobachtung von schädlichen Auswirkungen beim Einsatz von KI-Algorithmen in verschiedenen

Anwendungen führte zu einem neuen Forschungsgebiet namens Responsible AI, das sich mit den ethischen Fragestellungen befasst. Eine wichtige Komponente ist die Explainable AI: sie soll nachvollziehbare Erklärungen für die KI-Schlussfolgerungen liefern. Dies ist v.a. für die automatisierte Diagnostik und Entscheidungsunterstützung bei Clinical Decision Support Systems (siehe HCM 6/2018) nötig. Im Blog unter [www.serapion.de](http://www.serapion.de) stellt die KCC-Expertin Dr. Christina Czeschik ein Erklärungstool für KI-Entscheidungen namens Local Interpretable Model-Agnostic Explanations (LIME) vor.

Manfred Kindler, KCC-Vorsitzender,  
Kontakt: [m.kindler@kcc.info](mailto:m.kindler@kcc.info)

## KI-Software als Medizinprodukt

Nach einer dreijährigen Übergangszeit gilt ab 26. Mai 2020 die EU-Verordnung für Medizinprodukte (MDR). Sie hat gravierende Konsequenzen für die Entwicklung und Klassifizierung interpretativer Software. Die Änderung der Regel 11 für die Klassifizierung wird durch die notwendige Zertifizierung durch eine Benannte Stelle die Kosten und Aufwände v.a. für kleinere Hersteller, App-Entwickler und Start-ups vervielfachen. Sie besagt: „Software, die dazu bestimmt ist, Informationen zu liefern, die zu Entscheidungen für diagnostische oder therapeutische Zwecke herangezogen werden, gehört zur Klasse IIa, es sei denn, diese Entscheidungen haben Auswirkungen, die Folgendes verursachen können:

- den Tod oder eine irreversible Verschlechterung des Gesundheitszustands einer Person; in diesem Fall wird sie der Klasse III zugeordnet, oder;
- eine schwerwiegende Verschlechterung des Gesundheitszustands einer Person oder einen chirurgischen Eingriff; in diesem Fall wird sie der Klasse IIb zugeordnet.“