oder Empfehlung gelangt. Der Algorithmus ist eine sogenannte Black Box, deren Innenleben dem Nutzer verborgen bleibt. Menschliche Fachpersonen dagegen können darüber Auskunft geben, welche Informationen ausschlaggebend für eine Entscheidung waren – warum beispielsweise eine Zellveränderung im mikroskopischen Bild als wahrscheinlich bösartig eingestuft wird. Diese Divergenz ist ein Grund dafür, dass KI-Verfahren im Gesundheitswesen bisher nur geringes Vertrauen genießen und sich insgesamt weniger durchgesetzt haben, als dies in früheren Jahren angenommen wurde.

Ein Beispiel dafür ist IBM Watson bzw. dessen Anwendung Watson for Oncology. Eine Studie hat gezeigt, dass diese für die Krebsmedizin entwickelte KI nur dann das Vertrauen ihrer Nutzer (der Onkologen) genoss, wenn sie zu einer übereinstimmenden Diagnose gelangt waren. In diesem Fall wurde das Votum der KI als Bestätigung der eigenen Diagnose akzeptiert. Wenn sich die Ergebnisse der menschlichen Fachpersonen und von Watson unterschieden, hielten diese den Output von Watson für fehlerhaft, nicht die eigene Diagnose (Bloomberg 2018).

Hinzu kommt: Ärztinnen und Ärzte müssen ihre Entscheidungen auch gegenüber Patient:innen begründen und erklären. Dies stellt die Fachleute vor die Frage, wie sie eine Diagnose oder Therapieentscheidung gegenüber dem Betreffenden rechtfertigen können, wenn sie nicht vollständig nachvollziehen können, wie diese zustande gekommen ist.

Der Wunsch nach mehr Transparenz in der KI ist durchaus begründet. Bias (Voreingenommenheit) von Algorithmen als Fehlerquelle entsteht beispielsweise durch unausgewogene Zusammensetzung von Trainingsdaten (den Daten, anhand derer ein KI-Algorithmus lernt, korrekte Entscheidungen oder Klassifikationen zu treffen). So arbeiten etwa diagnostische Algorithmen in der Dermatologie fehlerhaft bei der Diagnose von Hautkrebs bei dunkelhäutigen Patient:innen, wenn sie mit Bildern von mehrheitlich hellhäutigen Patient:innen trainiert worden sind (Lashbrook 2018).

1.2.1 Responsible AI und Explainable AI

Die potenziell fatalen Auswirkungen des Einsatzes von KI in verschiedenen Fachbereichen und Wege, um diese zu vermeiden, werden seit einigen Jahren im einem Forschungsgebiet namens „Responsible AI“ („Verantwortungsbewusste KI“) untersucht. Das Forschungsfeld der „Explainable Artificial Intelligence“ (XAI, „Erklärbare KI“) beschäftigt sich ergänzend mit der Frage, wie Algorithmen menschlich verständliche Begründungen für ihre Entscheidungen liefern und damit die Transparenz erhöhen können.

Erste Untersuchungen (Hosanagar u. Jair 2018) zeigen, dass das Vertrauen der Nutzenden in ein unerwartetes Ergebnis durch eine Erklärung dieses Ergebnisses gesteigert werden kann. Auch die europäische Datenschutz-Grundverordnung (DSGVO) verlangt im Prinzip nach solchen Verfahren: Sie enthält ein Transparenzgebot, demzufolge bei einer automatisiert getroffenen Klassifikation oder Entscheidung Betroffene das Recht haben, Auskunft über die „bedeutungsvollen Parameter“ dieser Entscheidung zu erhalten.