Abstract

Die Integration von Künstlicher Intelligenz (KI) in medizinische Entscheidungsfindung bietet beispiellose Möglichkeiten zur Verbesserung der diagnostischen Genauigkeit und Entscheidungsunterstützung. Ihre undurchsichtige „Black-Box“-Natur stellt jedoch erhebliche Herausforderungen dar, insbesondere in Bezug auf das Vertrauen von Nutzern wie medizinischen Fachkräften und Patienten. Erklärbarkeit ist ein entscheidender Faktor zur Bewältigung dieser Herausforderungen, doch ihre Auswirkungen bleiben ambivalent. Diese Dissertation untersucht kritisch die positiven und negativen Auswirkungen von KI-Erklärungen im medizinischen Kontext und bietet eine differenzierte Perspektive auf deren Rolle in KI-gestützten medizinischen Entscheidungen.

Die Dissertation umfasst eine konzeptionelle und drei empirische Studien, die untersuchen, wie Erklärungen medizinische Fachkräfte und Patienten beeinflussen. Die empirischen Ergebnisse zeigen, dass Erklärungen die wahrgenommene Transparenz, Nützlichkeit und das kausale Verständnis erhöhen, was die Akzeptanz und das Vertrauen in KI-Systeme fördert. Sie wirken zudem wie ein „Impfstoff“ gegen die Abkehr von der Nutzung, indem sie das Vertrauen der Nutzer auch bei KI-Fehlern aufrechterhalten. Darüber hinaus verringern Erklärungen für einige Nutzer Datenschutzbedenken, indem sie Unsicherheiten über die Datennutzung reduzieren. Die Dissertation deckt jedoch auch erhebliche negative Effekte auf. Erklärungen können unbeabsichtigt die kognitive Belastung erhöhen, Nutzer in hochkritischen Situationen überfordern und Datenschutzbedenken verstärken, indem sie auf sensible Datenverarbeitung aufmerksam machen. Diese Ergebnisse unterstreichen den ambivalenten Charakter von Erklärungen und die Notwendigkeit einer sorgfältigen Implementierung, um Vorteile zu maximieren und Risiken zu minimieren. Zur Überbrückung von Theorie und Praxis entwickelt diese Dissertation ein prozedurales Modell, das die Integration von Erklärungen in die Phasen vor der Nutzung, während der Nutzung und nach der Nutzung leitet. Das Modell zeigt seine Nützlichkeit in hypothetischen Szenarien, wie etwa klinischen Entscheidungsunterstützungssystemen für Radiologen und Symptomprüfungsanwendungen für Patienten, und bietet maßgeschneiderte Strategien zur Balance zwischen Transparenz, kognitiver Belastung und Datenschutzüberlegungen.

Mit einem Beitrag zur Literatur über Mensch-KI-Interaktion und zur Diskussion über die Schattenseiten von Informationssystemen betont diese Forschung die strategische Bedeutung von Erklärungen für die langfristige Akzeptanz von KI-Technologien. Sie liefert theoretische Erkenntnisse, empirische Belege und umsetzbare Empfehlungen für die Gestaltung erklärbarer KI-Systeme, die den Bedürfnissen der Nutzer und regulatorischen Anforderungen entsprechen.