



Elektrotechnisches Kolloquium

der Bergischen Universität Wuppertal

Die Fakultät für Elektrotechnik, Informationstechnik und Medientechnik lädt zur Teilnahme an folgender Vortragsveranstaltung mit anschließender Diskussion ein:

Es spricht: Richard Meyes, M.Sc.

Lehrstuhl für Technologien und Management der digitalen Transformation

Prof. Dr.-Ing. Tobias Meisen

über das Thema

Transparency and Interpretability for Learned Representations of Artificial Neural Networks

Inhalt:

Künstliche Intelligenz (KI) ist ein Konzept, dessen Bedeutung und Wahrnehmung sich in den letzten Jahrzehnten erheblich gewandelt hat. Was in den 1950er Jahren mit vereinzelt Forschungsanstrengungen stark theoretischer Natur begann, hat sich zu einem vollwertigen Forschungsgebiet der Neuzeit entwickelt mit dem Potential zur einer der wichtigsten technologischen Errungenschaften der Menschheit zu werden. Außerhalb der akademischen Forschung wurde KI zunächst vor allem in Film und Fiktion dargestellt, bevor sie für praktische Anwendungen in Industrie und Gesellschaft an Bedeutung gewann. Dieser Fortschritt ist unter anderem auf den digitalen Wandel der letzten Jahre zurückzuführen, in denen die Verfügbarkeit von Rechenressourcen zur Verarbeitung wachsender Datenmengen stark zugenommen hat und die Anwendung von KI in Form von künstlichen neuronalen Netzen als Black-Box-Modelle in großem Maßstab wirtschaftlich nutzbar gemacht hat.

Trotz der rasanten technologischen Fortschritte dieser KI-Blackbox-Modelle im zweiten Jahrzehnt des 21. Jahrhunderts bleiben einige wichtige Fragen über die Art ihrer Entscheidungsfindung unbeantwortet. Warum entscheidet eine KI so, wie sie es tut? Was ist die Grundlage für ihre Entscheidungen? Warum wurde diese Grundlage anhand der vorliegenden Daten als wichtig erkannt? Inwieweit ist diese Grundlage mit menschlichen Vorstellungen von Moral und Ethik vereinbar und wie kann die Entscheidungsfindung einer KI bewusst beeinflusst werden, um diese Vereinbarkeit zu gewährleisten? All diese Fragen drehen sich um das Thema von Transparenz, Interpretierbarkeit und Erklärbarkeit der Entscheidungsfindung einer KI, ein junges Forschungsgebiet, das unter dem Oberbegriff XAI zusammengefasst wird und sich aus den immer strengeren Anforderungen an den Einsatz von KI in sicherheitskritischen oder ethisch sensiblen Bereichen ergibt.

Der Vortrag behandelt Methoden zur empirischen Untersuchung von KI-Blackbox-Modellen und deren erlernte Wissensrepräsentationen. Dabei stehen insbesondere durch den Forschungsbereich der Neurowissenschaften inspirierte Ansätze im Fokus, die für die Untersuchung von künstlichen neuronalen Netzen adaptiert wurden.

Termin: 01.06.2022, 14:00 Uhr

Ort: Bergische Universität Wuppertal
Campus Freudenberg, Dekanatssaal FME 1. Etage