13. Karlsruher Tag der IT-Sicherheit

20.07.2023 | 14:00 – 18:40 Uhr

15:05 Uhr | Sicherheit und Erklärbarkeit von KI

Künstliche Intelligenz (KI) ist in Form von Expertensystemen bereits seit mehreren Jahren fest in unserem Alltag verankert und die Verbreitung lernender Systeme nimmt stetig zu. Dazu gehören Empfehlungen beim Online-Shopping, Autopiloten beim autonomen Fahren oder auch Sprachassistenten. Moderne Lernverfahren, wie z. B. tiefe neuronale Netze (Deep Learning), bieten große Chancen für die breite Anwendung, eröffnen aber gleichzeitig eine neuartige Angriffsfläche: Durch geschickt manipulierte Eingabedaten kann die Entscheidung von KI beeinflusst werden.

Gleichzeitig ist es denkbar, dass die KI von einem Angreifer so manipuliert wird, dass sie in sehr spezifischen Fällen falsche Aussagen trifft. Hier kann es helfen die konkrete Funktionsweise von KI und warum bestimmte Entscheidungen getroffen wurden nachvollziehbar zu machen.

Methoden der erklärbaren künstlichen Intelligenz („XAI“) bieten diese Möglichkeit, müssen aber erst noch zeigen, dass sie sich für den robusten Einsatz eignen. Der Vortrag gibt einen Überblick über dieses spannende Forschungsfeld.

Professor Dr. Christian Wressnegger, KASTEL*, ist Professor für Informatik am Karlsruher Institut für Technologie (KIT) und leitet die Forschungsgruppe "Intelligente Systemsicherheit". Er forscht an der Schnittstelle des maschinellen Lernens und der IT-Sicherheit. Dabei beschäftigt er sich sowohl mit dem Einsatz von KI zum Schutz von Computersystemen als auch mit der Sicherheit und Erklärbarkeit von KI.*

www.tag-der-it-sicherheit.de

Profitieren auch Sie von den Erfahrungen anderer - wir freuen uns auf Ihr Kommen!