

Press release

SRH University

Marie-Luise Unteutsch

02/06/2025

<http://idw-online.de/en/news847035>Research projects, Scientific Publications
Information technology
transregional, national

SRH University erforscht Einsatz von erklärbarer künstlicher Intelligenz zur Erkennung von Deepfakes

Eine neue Studie der SRH University unterstreicht den Nutzen von erklärbaren KI-Systemen zur sicheren und transparenten Erkennung von Deepfakes. KI-Entscheidungen können durch Merkmalsanalysen und Visualisierungen nachvollziehbar dargestellt werden und so das Vertrauen in KI-Technologien fördern.

Ein Forscherteam um Prof. Dr. Alexander I. Iliev von der SRH University, mit wesentlichen Beiträgen des Forschers Nazneen Mansoor, hat eine innovative Methode zur Erkennung von Deepfakes entwickelt. In der jüngst in der Fachzeitschrift Applied Sciences veröffentlichten Studie präsentieren die Wissenschaftler den Einsatz von erklärbarer künstlicher Intelligenz (Explainable AI), um die Transparenz und Zuverlässigkeit bei der Identifizierung manipulierter Medieninhalte zu erhöhen.

Deepfakes, also mittels künstlicher Intelligenz erzeugte gefälschte Medieninhalte wie Videos oder Audiodateien, stellen eine zunehmende Bedrohung für die Gesellschaft dar, da sie zur Verbreitung von Fehlinformationen und zur Beeinträchtigung des öffentlichen Vertrauens genutzt werden können. Die herkömmlichen Erkennungsmethoden stoßen dabei oft an ihre Grenzen, insbesondere wenn es darum geht, die Entscheidungsprozesse der KI-Modelle nachvollziehbar zu machen.

Das Team der SRH University führte in seiner Studie umfangreiche Tests durch, bei denen unterschiedliche KI-Modelle auf ihre Fähigkeit geprüft wurden, Deepfakes zuverlässig zu identifizieren. Besonderes Augenmerk lag dabei auf erklärbarer KI, die es ermöglicht, die Entscheidungsgrundlage der Modelle transparent und nachvollziehbar darzustellen. Dies geschieht beispielsweise durch Visualisierungstechniken wie „Heatmaps“, die farblich markieren, welche Bildbereiche die KI als relevant für ihre Entscheidung identifiziert hat. Darüber hinaus analysieren die erklärbaren Modelle spezifische Merkmale, wie Texturen oder Bewegungsmuster, die auf eine Manipulation hinweisen.

Prof. Dr. Alexander I. Iliev, Studiengangsleiter im Master-Studiengang Computer Science – Big Data & Artificial Intelligence, erläutert die Bedeutung dieser Ansätze: „Unser Ziel war es, Technologien zu schaffen, die nicht nur effektiv, sondern auch vertrauenswürdig sind. Die Fähigkeit, den Entscheidungsprozess der KI transparent zu machen, wird zunehmend wichtiger – sei es in der Strafverfolgung, der Medienbranche oder in der Wissenschaft.“

Die Studie zeigt, dass erklärbare KI nicht nur die Erkennungsgenauigkeit verbessert, sondern auch das Verständnis und Vertrauen in KI-Technologien fördert. Indem sie aufzeigt, wie die Entscheidungen getroffen wurden, können Schwachstellen in den Modellen identifiziert und zukünftige Systeme gezielt optimiert werden. Dies ist ein entscheidender Schritt, um die verantwortungsvolle Nutzung von KI in der Gesellschaft zu stärken.

Die SRH University unterstreicht mit dieser Forschung ihre führende Rolle im Bereich der angewandten Wissenschaften und der Entwicklung innovativer Technologien. Das Studienangebot der Hochschule, wie beispielsweise die Masterstudiengänge Computer Science oder Information Technology, bereitet Studierende gezielt auf aktuelle

Herausforderungen im Bereich der künstlichen Intelligenz vor.

Die vollständige Studie mit dem Titel "Explainable AI for DeepFake Detection" ist in der Fachzeitschrift Applied Sciences veröffentlicht und unter folgendem Link abrufbar: <https://www.mdpi.com/2076-3417/15/2/725>

contact for scientific information:

Prof. Dr. Alexander I. Iliev
alexander.iliev@srh.de