

Täuschend echt: KI-generierte Histologiebilder

Selbst Experten können künstlich erzeugte Gewebeschnittbilder nicht zuverlässig identifizieren / Herausforderung für die Beurteilung biomedizinischer Forschung

Sie hilft bei der Beurteilung von Röntgen- und MRT-Bildern, beim Schreiben von Arztbriefen und bei der Auswertung von Messdaten in der Forschung – die künstliche Intelligenz, kurz KI, ist in der klinischen Praxis und in der biomedizinischen Forschung zu einem hilfreichen und mächtigen Instrument geworden. So mächtig, dass Hochschulen und Forschungsorganisationen Empfehlungen und Leitlinien zu ihrer Nutzung in der Wissenschaft formulieren.

Wie berechtigt das ist, untermauert das Experiment einer Arbeitsgruppe am Universitätsklinikum Jena: Das Forschungsteam nutzte den öffentlich verfügbaren Deep-Learning-Algorithmus Stable Diffusion, um aus Trainingsdaten neue histologische Bilder zu erstellen. „Wir verwendeten als Trainingsbilder einmal drei und einmal 15 echte Schnittbilder von angefärbtem Gewebe aus Mäusenieren“, so Studienleiter Prof. Ralf Mrowka. „Bereits auf der Basis von drei Originalbildern entstanden schon recht überzeugende Bilder.“

In einer Online-Umfrage testete das Forschungsteam, ob diese KI-generierten Histologiebilder von echten zu unterscheiden waren. Die über 800 Teilnehmenden der Umfrage wurden dafür in Experten oder Laien eingeteilt, je nachdem, ob sie Erfahrungen mit dem Beurteilen von histologischen Bildern hatten, wie etwa Medizinstudierende, oder nicht. Ihnen wurden nacheinander 16 einzelne Bilder – je acht echte und KI-generierte – gezeigt, jedes Bild sollte klassifiziert werden, bevor das nächste erschien.

Neue Methoden zur Erkennung gefälschter Daten in wissenschaftlichen Arbeiten notwendig

Die Expertengruppe ordnete gut zwei Drittel richtig ein. Der Laiengruppe gelang dies nur in gut der Hälfte der Fälle. Beide Gruppen konnten die auf weniger Trainingsdaten beruhenden KI-Bilder häufiger enttarnen. Zehn Teilnehmende der Expertengruppe erkannten alle Bilder korrekt. Über alle Bilder- und Teilnehmergruppen hinweg fielen die Entscheidungen zumeist innerhalb der ersten halben Minute. „Wir konnten auch feststellen, dass richtige Zuordnungen signifikant schneller getroffen wurden als falsche,“ stellt Erstautor Dr. Jan Hartung heraus, „eine Beobachtung, die mit gängigen Modellen der wahrnehmungsbasierten Entscheidungsfindung im Einklang steht“.

Ralf Mrowka: „Unser Experiment zeigt, dass die Erfahrung hilft, gefälschte Bilder zu erkennen; dass aber auch dann ein nicht geringer Anteil künstlicher Bilder nicht zuverlässig identifiziert wird.“ „Dabei sind die technischen Hürden für jemanden mit der Absicht, Abbildungen zu fälschen, relativ gering“, betont Jan Hartung. „Man braucht nicht hunderte von Beispielen zum Trainieren eines Algorithmus, ein Dutzend kann bereits ausreichend sein.“ Das Autorenteam sieht darin eine Herausforderung für die Wissenschaft, derer diese sich bewusst sein muss und die neuer Methoden zur Erkennung gefälschter Daten in wissenschaftlichen Arbeiten bedarf.

Originalpublikation:

Hartung, J., Reuter, S., Kulow, V.A. et al. Experts fail to reliably detect AI-generated histological

data. Sci Rep 14, 28677 (2024). <https://doi.org/10.1038/s41598-024-73913-8>