

KI braucht ein Etikett

15.6.2026 - | Friedrich-Alexander-Universität Erlangen-Nürnberg

FAU-Informatiker Christian Riess erklärt, warum die Kennzeichnung KI-generierter Inhalte so schwierig ist. Ein Foto, das nie aufgenommen wurde. Ein Text, den kein Mensch geschrieben hat. Beides gehört heute zum Alltag, doch der Umgang mit KI-generierten Inhalten berührt Grundfragen von Demokratie, Medienvertrauen und digitaler Öffentlichkeit. Im EU AI Act will Europa unter anderem regeln, wie solche Inhalte künftig sinnvoll und rechtssicher erkennbar gemacht werden sollen. Mit dabei: Informatiker Prof. Dr. Christian Riess vom Lehrstuhl für Angewandte Kryptographie an der Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU). Er hat in einer Gruppe von Expert/-innen Handlungsempfehlungen - einen Code of Practice - für die Kennzeichnung von KI-Inhalten erarbeitet.

Viele fordern eine Kennzeichnung von KI-generierten Inhalten. Klingt einfach - warum ist die Umsetzung in der Praxis so kompliziert?

Schwierig ist es, mit einer einzigen Rahmen-Regelung der Vielfalt der möglichen Medien und Anwendungsfälle gerecht zu werden, so dass das eigentliche Ziel erreicht wird, nämlich Nutzer/-innen zu informieren, wenn sie auf KI-Inhalten treffen.

Wir unterscheiden Kennzeichnungen, die für Menschen wahrnehmbar sind und Kennzeichnungen, die technisch ausgelesen werden. Eine geeignete Form von wahrnehmbaren Kennzeichen hängt immer vom Medium und Anwendungsfall ab, z.B. bei einem Tondokument ein automatischer Einspieler der Art „Dieses Tondokument wurde KI-generiert“ oder eine kleine „AI“-Markierung am Rand eines Bilds.

Nicht-wahrnehmbare Kennzeichnungen sind wichtig, um automatisiert die Information, dass ein Dokument KI-generiert ist, auszulesen und den Nutzer/-innen kenntlich zu machen. Hier ist die Herausforderung, dass die Kennzeichnung die Verarbeitungsschritte der Plattformen [wo die Daten schlussendlich veröffentlicht werden] überlebt. Es gibt beispielsweise standardisierte Verfahren, um in den Metadaten von Dokumenten Kennzeichnungen einzufügen. Aber große Plattformen entfernen solche Metadaten automatisch von Dokumenten. Daher fordert der Code of Practice, dass die Kennzeichnung in Form eines digitalen Wasserzeichens direkt mit dem Inhalt verwoben wird und bei Veröffentlichung auf den Plattformen ausgelesen werden kann. Im Detail gibt es hier noch viele weitere technische Fragestellungen, die gelöst werden müssen.

Am Verhandlungstisch sitzen Unternehmen, Zivilgesellschaft und Rechtsexpert/-innen mit teils sehr unterschiedlichen Vorstellungen. Wo prallen die Interessen am stärksten aufeinander?

Grundsätzlich ist jede Form von Regulierung eine Belastung für Unternehmen, demgegenüber ein berechtigtes Interesse der Zivilgesellschaft nach Auskunft über Interaktion mit KI-generierten Inhalten steht. Aus Sicht der Zivilgesellschaft ist es wichtig, eine Lösung zu finden, die für die Bürgerinnen und Bürger Europas gut funktioniert, ohne dabei mögliche private Informationen von Einzelpersonen preiszugeben. Aus Sicht der Unternehmen ist es wichtig, Lösungen zu finden, die einerseits kostengünstig umsetzbar sind und die andererseits den eigentlichen KI-generierten Inhalt nicht so weit beeinträchtigen, dass er nicht mehr den intendierten Zweck erfüllt.

Und woran erkennt man, dass ein tragfähiger Kompromiss gefunden ist?

Unternehmen können als Unterzeichner direkt ihre Unterstützung für den Code of Practice demonstrieren und so gleichzeitig Rechtssicherheit erlangen. Die Zivilgesellschaft kann sich darauf verlassen, dass die Unterzeichner künstlich generierte Inhalte gemäß den formulierten Standards kennzeichnen. Für die Zivilgesellschaft und insbesondere Berufsgruppen wie Journalist/-innen, Fact Checker/-innen, Wissenschaftler/-innen und Strafverfolgungsbehörden, wird sich die Qualität der Regelungen in der konkreten Arbeit mit den Detektoren für die Markierungen zeigen. Für beide Parteien ist es wichtig zu wissen, dass der Code of Practice aktualisiert werden kann, wenn sich herausstellen sollte, dass bestimmte Elemente nicht praktikabel sind.

Wo verläuft die Grenze zwischen menschlicher Arbeit und KI-generiertem Inhalt? Müssen Studierende ihre Arbeiten jetzt als KI-Inhalte kennzeichnen, wenn ChatGPT nur einzelne Formulierungen vorgeschlagen hat?

Wir unterscheiden Markierungen für vollständig KI-generierte Inhalte, Markierungen für KI-modifizierte Inhalte und KI-Modifikationen die lediglich unterstützende Funktion haben wie z.B. eine Rechtschreibprüfung. Die Grenzen sind aber tatsächlich fließend und Bildungseinrichtungen wie auch die FAU arbeiten richtigerweise eigene, spezifische Richtlinien heraus welche Form von KI-Unterstützung in welchem Kontext angemessen ist.

Woran werden wir perspektivisch erkennen, dass die Regulierung funktioniert hat?

Die intendierte gesellschaftliche Wirkung zur besseren Bekämpfung von Desinformation muss im Nachgang an die Implementierung wissenschaftlich untersucht werden. Der Code of Practice kann jedoch nicht über die Regelungen der KI-Verordnung hinausgehen, so ist es beispielsweise im Rahmen der KI-Verordnung nicht möglich, eine explizite Markierung menschlich erzeugter (nicht-KI-)Inhalte zu fordern.

<https://www.fau.de/2026/06/news/ki-braucht-ein-etikett>