

# Die Kalibrierung des Vertrauens

Viele Menschen suchen nach neuen Berufen oder sind unsicher bei der Studienwahl. KI kann dabei helfen, Begabungen frühzeitig zu erkennen und Fehleinschätzungen zu vermeiden. Dafür braucht es Vertrauen in die Technologie. Das entsteht u.a. durch gute Leistung, nachvollziehbare Prozesse und verlässliche Visualisierungen, z.B. vom Konfidenzniveau.

VON PROF. DR. SABRINA SCHORK

**DIE AKTUELLEN HERAUSFORDERUNGEN** im Bildungsbereich werden durch Studien deutlich. Laut Gallup-Institut 2023 sind viele deutsche Arbeitnehmer auf der Suche nach neuen Möglichkeiten. Ein Drittel der Studierenden bricht das Studium ab oder wechselt die Fachrichtung (DZHW-Bericht 2024). Dies könnte daran liegen, dass einige Menschen nicht wissen, worin ihr Potenzial liegt oder was sie wirklich wollen, so die These von Professorin Schork.

Für Lehrkräfte ist es schwierig, Begabungen zu erkennen, und nur wenige sehen es als ihre Aufgabe an, Begabungen zu erkennen und zu fördern. Zudem haben viele Menschen eine Abneigung gegenüber (Hoch-)Begabten und grenzen sie aus. Medizinisches Fachpersonal ist selten in der Lage, Hochbegabung und Hochsensibilität zu erkennen, was zu Fehldiagnosen und Fehlbehandlungen führen kann (Germann-Tillmann & Treier 2021; Webb & Latimer 1983). Diese Fehleinschätzungen haben weitreichende ökonomische und soziale Folgen. Verkannte Genies können unglücklich und krank werden, anstatt

EU ist jedoch stark reguliert (EU 2024) und Institutionen eher veränderungsträge. Vertrauen vermittelt zwischen Mensch und KI im Rahmen der Lernenden-Lehrenden-Beziehung.

Damit Vertrauen in KI auf unterschiedlichen Ebenen (persönlich, institutionell) im Bildungsbereich entstehen kann, müssen verschiedene Faktoren erfüllt sein:

- Genauigkeit, Glaubwürdigkeit, Aktualität, Präzision, Relevanz und Zuverlässigkeit der Informationen, die eine KI an den Menschen kommuniziert, erhöhen das Vertrauen in die KI (Chi et al. 2023).
- Vertrauen in die Technologie entsteht durch ihre Leistung (was die Automatisierung tut), durch die Nachvollziehbarkeit ihrer Funktionslogik (Prozessinformationen darüber, warum eine Maschine eine Entscheidung getroffen hat) und durch ihre Absichten (warum die Automatisierung entwickelt wurde). Vertrauen entsteht zudem durch Auseinandersetzung mit der KI. Standards können dabei einen rechtssicheren Rahmen bieten.



## DIE AUTORIN

**Prof. Dr. Sabrina Schork**

ist Forschungsprofessorin an der Technischen Hochschule Aschaffenburg und Autorin des Buches „Vertrauen in Künstliche Intelligenz“.




gesellschaftliche Herausforderungen zu lösen, wozu sie besonders befähigt sind (Schork 2024).

KI kann dabei unterstützen Begabungen frühzeitig zu erkennen und ist daher ein Hoffnungsträger für die persönliche Entwicklung junger Erwachsener (Schleiss et al. 2023). Es stellt sich die Frage, warum KI nicht genutzt werden sollte, um das Potenzial und die Begabungen von Lernenden zu erschließen. Dies erfordert das Vertrauen des Staates, der Lehrenden und der Lernenden in die KI und die Menschen dahinter. Das Bildungswesen in der

Datensicherheit ist entscheidend, und die Datengebenen sollten wissen, was mit ihren Daten passiert. Der Algorithmus sollte fehlerfrei arbeiten, und der Mensch sollte die Kontrolle behalten und Entscheidungen nachvollziehen können. Die Einbeziehung der Nutzenden in die Entwicklung der KI fördert Vertrauen (Hengstler et al. 2016).

- In einem Experiment wurde untersucht, wie kalibriertes Vertrauen zwischen Menschen und KI entstehen kann.



KI kann dabei unterstützen Begabungen frühzeitig zu erkennen und ist daher ein **Hoffnungsträger** für die persönliche Entwicklung junger Erwachsener.

Die Erklärung der KI-Entscheidungen führte zu einem ausgeglichenen Vertrauensniveau bei den Menschen. Visuelle Erklärungen der KI-Entscheidungen wirkten sich positiv auf das menschliche Verhalten und die Vertrauenskalibrierung aus. Im Gegensatz dazu hatten einfache Aufklärungsmaßnahmen mit klassischen Text-Bild-Kombinationen und domänenspezifischem Wissen keinen Einfluss auf das Vertrauen in die KI. Obwohl visuelle Erklärungen kurzfristig zu negativem Nutzerfeedback führen können, wird die KI langfristig an ihrer tatsächlichen Leistung gemessen. Fehleinschätzungen der Technologie können einschneidende Folgen haben. Fehler von KI sollten daher aus ethischer und ökonomischer Sicht transparent gemacht werden (Leichtmann et al. 2023).

- Die vier Kategorien von Technologien des maschinellen Lernens – Fairness, Erklärbarkeit, Überprüfbarkeit und Sicherheit (FEAS) – müssen zur Vertrauensbildung in KI in den verschiedenen Phasen des Lebenszyklus eines Systems berücksichtigt werden. Jede dieser Phasen ist Teil einer Vertrauensketten (Toreini et al. 2020).
- Falls Expertenentscheidungen durch KI gestützt werden sollen, liegt der Schlüssel darin, das Vertrauen des Menschen in die KI von Fall zu Fall angemessen zu kalibrieren. Die Angabe des Konfidenzniveaus kann dabei helfen. Die Kalibrierung des Vertrauens allein reicht jedoch nicht aus, um die KI-gestützte Entscheidungsfindung zu verbessern. Dies hängt auch davon ab, ob Menschen genügend einzigartiges Wissen einbringen können, um die Fehler der KI zu ergänzen (Zhang et al. 2020).

- Die Darstellung der KI (robotisch, virtuell, eingebettet) und ihrer Fähigkeiten sind wichtige Faktoren für die Vertrauensbildung. Kognitives Vertrauen in KI wird durch Greifbarkeit, Transparenz, Zuverlässigkeit und Unmittelbarkeit gestärkt. Und emotionales Vertrauen in KI entsteht durch ihre Menschenähnlichkeit. Unmittelbares Verhalten kann kognitives Vertrauen schaffen, aber emotionales Unbehagen hervorrufen. Nutzerinnen und Nutzer vertrauen einer KI eher, wenn es um technischen Rat geht, und einem Menschen eher, wenn es um soziale Fragen geht. Darüber hinaus sind Nutzende gegenüber KI ehrlicher und offener als gegenüber Menschen (Glikson & Williams-Woolley 2020).

Um KI im Bildungsbereich zu etablieren, ist es unerlässlich, gemeinsam mit den Lernenden verlässliche, kontrollierbare und nachvollziehbare KI-Systeme nach den FEAS-Prinzipien zu entwickeln und qualitativ hochwertige Mensch-Maschine-Dialoge zu etablieren, in denen sich junge Erwachsene öffnen können und gleichzeitig sicher sein können, dass die geteilten Informationen nicht weitergegeben werden (was bei Menschen nicht der Fall ist).

KI kann im Bildungskontext also die Mensch-Mensch-Beziehung bereichern und nur dann erfolgreich sein, wenn die Menschen verstehen, wie sie funktioniert, und in der Lage sind, den Grad des Vertrauens in der Interaktion mit ihr zu kalibrieren. Der Unterschied zwischen Menschen und Maschine soll erhalten bleiben und sich gegenseitig positiv verstärken. KI kann dabei unterstützen Potenziale frühzeitig zu erkennen und diese Begabungen in Leistungsstärke zu überführen. Datengestützte Erkenntnisse und personalisierte Beratung leisten einen wichtigen Beitrag zur Entfaltung individueller Kräfte. •