

Wie die Notengebung durch die Künstliche Intelligenz gerechter werden kann

Wer hat eine Note nicht schon einmal als ungerecht und beliebig empfunden? Die Internationale Hochschule (IU) erforscht die Möglichkeit eines Systems für eine gerechtere Notenvergabe mit Hilfe der Künstlichen Intelligenz.

[Dr. Sven Schütt](#)

08.01.2025 Lesezeit 3 min

Jede Note fasst unzählige Einzelleistungen in einer einzigen Zahl zusammen, und wir alle haben eine Note schon einmal als ungerecht und beliebig empfunden. Da diese traditionellen Noten über den weiteren Ausbildungs- und Karriereweg entscheiden, hat die Internationale Hochschule (IU) erforscht, wie man ein System entwickeln kann, um Noten durch KI gerechter zu vergeben.

Die Notengebung spiegelt heute im Wesentlichen die Rangfolge innerhalb einer spezifischen Lerngruppe wider, wie etwa in einem Seminar, das sich in seiner Zusammensetzung von anderen Gruppen unterscheidet. Dadurch kann eine Note in einer Kohorte eine andere Leistung repräsentieren als in einer anderen Lerngemeinschaft. Die Noten sind nicht übergreifend vergleichbar. Außerdem sind Noten auch schon in den einzelnen Lerngruppen per se nicht objektiv: Die Leistung eines Lernenden setzt sich aus vielen Facetten zusammen – Präsentationen, Klausuren, persönliche Entwicklungsgespräche etc.

Lebenswege gestalten

Das grundlegende Problem liegt darin, dass eine Note eine subjektiv geprägte Zusammenfassung darstellt, die nicht wirklich zeigt, was eine Person tatsächlich kann oder wo ihre Stärken in einem bestimmten Schul- oder Studienfach liegen. Eine Ausnahme stellen sicherlich die naturwissenschaftlichen Fächer dar, die deutlich weniger Beurteilungsspielraum lassen, weil es hier mehr um „wahr oder falsch“ geht.

Eine Lösung wären sogenannte Rasterzeugnisse, welche einzelne Fähigkeiten auflisten und durch den Facettenreichtum zu einer gerechteren Notengebung beitragen. Dabei wird transparenter, was ein Mensch kann. Und das kann sogar zur gezielteren Gestaltung des weiteren Lebensweges beitragen.

Doch auch wenn die Notengebung grundsätzlich detaillierter und damit aufwändiger werden würde, bleibt ein gewisses Maß an Subjektivität bestehen. Lehrkräfte bewerten nach individuell geprägten Vorstellungen von Gerechtigkeit, die je nach historischem, soziokulturellem oder nationalem Kontext unterschiedlich ausfallen können. Manche setzen verstärkt auf rechnerisch zusammengesetzte Noten, andere vergeben sie nach einem Diskurs mit anderen Lehrenden oder den Lernenden. Je nachdem, wie die einzelnen Überzeugungen gewichtet sind und wie eng die Vorgaben der regionalen Gesetzgebung sind, werden Noten unterschiedlich vergeben. Es ist anzumerken, dass die Gerechtigkeitsüberzeugungen der Lehrkräfte oft nicht bewusst sind, da menschliche Urteile in der Regel von unbewussten Einstellungen beeinflusst werden.

KI – eine neue Hilfe für die Notengebung?

Nun kommt ein weiterer Spieler auf dieses weite Feld der Notengebung – die Künstliche Intelligenz,

kurz KI. Sie kann mit entsprechenden Prompts sehr viel leisten, beispielsweise Aufsätze zusammenfassen, prüfen, ob die richtigen inhaltlichen Aspekte erwähnt wurden, Rechtschreib- und Grammatikfehler erkennen und nach Vorgaben benoten. Um einen ersten Schritt zu gehen und zu eruieren, wie Künstliche Intelligenz die Lehrkräfte gezielt bei der Bewertung unterstützen kann, haben wir an der IU eine Untersuchung durchgeführt. Sie sollte herausfinden, ob Prüfungen von KI in großem Maßstab ausgewertet werden können und was die KI zu einer gerechteren Beurteilung beitragen kann. Ich greife kurz das Ergebnis vorweg: Mit dem Einsatz von KI werden Noten insgesamt gerechter, denn bei der Benotung von Prüfungen entfallen die individuellen Gerechtigkeitsüberzeugungen der Lehrkräfte.

Automatisierte Auswertung von Prüfungen

Für eine Forschungsarbeit hat ein Expertenteam der IU ein neuartiges KI-Modell mit einer enormen Menge an Prüfungsdatensätzen aus verschiedenen Fächern trainiert – von Geisteswissenschaften bis hin zu den MINT-Studiengängen. Das System „ASAG“, kurz für Automatic Short Answer Grading, kann kurze, individuell verfasste Antworten auf offene Fragen automatisiert bewerten.

Die Analyse belegt: Durch sein breites Spektrum an Trainingsdaten liefert das von der IU entwickelte ASAG-Modell brauchbare Ergebnisse für unterschiedlichste Disziplinen. Auch Antworten aus ihr bisher unbekanntem Themengebieten konnte die KI auswerten. Ein anschließender Vergleich mit Bewertungen von Lehrenden zeigt, dass ASAG bei der Bewertung der Antworten von Studierenden im Schnitt näher an der Referenznote lag, als dies bei der Bewertung der Lehrkräfte der Fall war (die Basis dafür waren bereits korrigierte Klausuren, die für den Vergleich vom KI-System und von Lehrenden erneut benotet wurden).

Die größere Übereinstimmung mit der Referenznote deutet darauf hin, dass eine KI-gestützte Benotung menschliche Subjektivität reduzieren und so potenziell die Fairness verbessern kann. Die Automatisierung der Benotung von Prüfungen mit offenen Fragen würde für Lehrende auf jeden Fall eine erhebliche Arbeitserleichterung darstellen und könnte dazu beitragen, menschliche Ungenauigkeiten und Fehler zu reduzieren.

Die KI wird auch das Bildungssystem verändern

Mit dem ASAG-Modell wollen wir die Möglichkeit schaffen und vorantreiben, KI zusätzlich zur menschlichen Benotung einzusetzen, um damit die Konsistenz und Fairness für die Studierenden zu erhöhen sowie negative Auswirkungen zu minimieren. Außerdem ist dieses Modell ein erster Schritt, um die eingangs erwähnten Benotungssysteme weiterzuentwickeln.

Nach Ansicht unseres Forschungsteams könnte das ASAG-Modell schon heute als Bewertungssystem eingesetzt werden. Die Prüfenden würden so eine Möglichkeit erhalten, die eigenen Bewertungen mit denen der KI abzugleichen. Dies würde zu verminderter Varianz sowie weniger Fehlern und erhöhter Fairness der Bewertung führen. Aktuell sind die rechtlichen und akademischen Voraussetzungen jedoch noch nicht gegeben. Entsprechende Rahmenbedingungen müssen noch geschaffen werden.