

## Künstliche Intelligenz: Trainingsdaten müssen gut, fair und ausgewogen sein

20. März 2023

Wie kann beim Einsatz Künstlicher Intelligenz (KI) gewährleistet werden, dass nicht nur die Datenqualität stimmt, sondern auch ethische und juristische Anforderungen erfüllt werden? In diese Problematik maschinellen Lernens führt ein kostenloser Onlinekurs ein, der am 19. April auf openHPI, der offenen Bildungsplattform des Hasso-Plattner-Instituts, startet. Geleitet wird er von HPI-Professor Felix Naumann und drei anderen Fachleuten: Medienethikerin Dr. Jessica Heesen von der Universität Tübingen, Strafrechts-Professorin Frauke Rostalski von der Universität Köln und Normungsexperte Dr. Sebastian Hallensleben vom Verband der Elektrotechnik, Elektronik und Informationstechnik. Anmelden für den zweiwöchigen Gratiskurs „KI und Datenqualität - Perspektiven aus Data Science, Ethik, Normung und Recht“ können sich alle Interessierten unter <https://open.hpi.de/courses/kidaten2023>.

Wer teilnimmt, bekommt vermittelt, wie unterschiedlich Expertinnen und Experten der Bereiche Informatik, Recht, Ethik und Normung auf die Fragestellung rund um solche Big-Data-Anwendungen blicken. „Wenn wir Künstliche Intelligenz gesellschaftlich verträglich nutzen wollen, benötigen wir für das Training der Modelle riesige Mengen an Daten, die passen und qualitativ hochwertig sind“, betont Naumann, der am HPI das Fachgebiet Informationssysteme leitet. Zusammen mit den anderen Fachleuten aus der Kursleitung hat er die vielfältigen Ansprüche an die Qualität „guter“ Daten im Blick.

### **„Schlechte Datenqualität kann zu Fehlentscheidungen führen“**

„Wir zeigen Einsteigern ins Thema, auf welche Aspekte bei der Datensammlung und Verarbeitung man achten sollte, um gute, faire und ausgewogene Trainingsdaten zu verwenden und so auch faire KI-Systeme zu entwickeln“, verspricht der HPI-Wissenschaftler. Nach seinen Worten wirken Anforderungen wie etwa Diskriminierungsfreiheit, Berücksichtigung von Diversität oder Arbeitnehmerdatenschutz auf die Daten und Prozesse zurück,

mit denen KI-Modelle zuvor trainiert wurden. „Umgekehrt führen unvollständige, fehlerbehaftete, unpassende oder einseitige Trainingsdaten zu unsicheren Modellen“, warnt Naumann. Die Ergebnisse könnten somit letztlich zu Fehlentscheidungen führen.

Zusammen mit den anderen Kursleitern will der Potsdamer Informatikwissenschaftler aufzeigen, dass auch die rechtlichen Vorgaben für Test-, Validierungs- und Trainingsdaten im maschinellen Lernen sowie deren Umsetzung in Normen und Standards noch „weitgehend ungeklärt“ sind. Die vier Dozenten des openHPI-Onlinekurses forschen im Rahmen des vom Bundesarbeitsministerium geförderten Projekts [KITQAR](#) gemeinsam an dem Thema KI und Datenqualität. Das Forschungskonsortium will bis Ende dieses Jahres praktisch anwendbare Qualitätsstandards für Test-, Validierungs- und Trainingsdaten im Bereich Künstliche Intelligenz entwickeln.

Für die Nutzung von Lehrvideos, Selbsttests, Hausaufgaben und Prüfungen sowie den Austausch im Kursforum sollten die Teilnehmenden einen Zeitaufwand von bis zu fünf Stunden pro Woche kalkulieren, rät Naumann. Besondere technische Vorkenntnisse bräuchten Interessierte nicht mitzubringen, denn relevante Grundlagen des maschinellen Lernens würden in der Einführung erläutert.

### **Hintergrund zur Bildungsplattform openHPI**

Seine interaktiven Kursangebote hat das Hasso-Plattner-Institut als Pionier unter den europäischen Wissenschafts-Institutionen am 5. September 2012 gestartet - auf der Internet-Plattform <https://open.hpi.de>. Diese bietet seitdem einen Gratis-Zugang zu aktuellem Hochschulwissen aus den sich schnell verändernden Gebieten der Informationstechnologie und Innovation. Das geschieht bislang hauptsächlich auf Deutsch und Englisch. Im Herbst 2017 hat openHPI aber erstmals auch die Online-Übersetzung und Untertitelung eines Kurses in elf Weltsprachen angeboten. Mittlerweile wurden auf openHPI fast 1,2 Millionen Kurseinschreibungen registriert. Rund 325.000 Personen aus 180 Ländern gehören derzeit auf der Plattform zum festen Nutzerkreis. Er wächst täglich. Für besonders erfolgreiche Teilnehmer an seinen "Massive Open Online Courses", kurz MOOCs genannt, stellte das Institut bisher rund 131.000 Zertifikate aus. Das openHPI-Jahresprogramm umfasst zahlreiche Angebote für IT-Einsteiger und Experten. Auch die in der Vergangenheit angebotenen gut 100 Kurse können im Selbststudium nach wie vor genutzt werden – ebenfalls kostenfrei. Studierende können sich für das Absolvieren von openHPI-Kursen jetzt auch Leistungspunkte an ihrer Universität anrechnen lassen. Wer sich Videolektionen aus den Kursen unterwegs auch dann anschauen will, wenn keine Internetverbindung gewährleistet ist (etwa im Flugzeug), kann zudem die openHPI-App für Android-Mobilgeräte, iPhones oder iPads nutzen. Partnerplattformen, die mit derselben Lerntechnologie arbeiten, sind neben [openSAP](#) und [OpenWHO](#) zum Beispiel auch [KI-Campus](#), [eGov-Campus](#) und [Kommunalcampus](#).

---

Pressekontakt: [presse@hpi.de](mailto:presse@hpi.de)

Christiane Rosenbach, Tel. 0331 5509-119, [christiane.rosenbach@hpi.de](mailto:christiane.rosenbach@hpi.de) und

Joana Bußmann, Tel. 0331 5509-375, [joana.bussmann@hpi.de](mailto:joana.bussmann@hpi.de)