

## Erforschung von ethischer KI



Stiftung Mercator vergibt 3,8 Millionen Euro für die Universitäten Bonn und Cambridge

Künstliche Intelligenz (KI) verändert die Gesellschaft tiefgreifend: Algorithmen beeinflussen den Zugang zu Berufen und Versicherungen, zur Justiz und medizinischen Behandlungen sowie unsere alltäglichen Interaktionen mit Freundinnen und Freunden sowie der Familie. Je schneller sich diese Technologien entwickeln, desto deutlicher werden die gesellschaftlichen Folgen ihres Einsatzes: Algorithmen, die rassistische Vorurteile im Gesundheitswesen fördern bis hin zur Verbreitung von Falschinformationen, die das Vertrauen in Demokratien untergraben. Um einen Beitrag zu leisten, dass KI grundlegende menschliche Werte unterstützt, fördert die Stiftung Mercator die Zusammenarbeit zwischen den Universitäten Bonn und Cambridge mit 3,8 Millionen Euro.

Unter Leitung von Prof. Markus Gabriel des Bonner Instituts für Philosophie und Dr. Stephen Cave vom Leverhulme Centre for the Future of Intelligence in Cambridge wird das Projekt Wünschenswerte Digitalisierung (Desirable Digitalisation: Rethinking AI for Just and Sustainable Futures) ethische Prinzipien in den Mittelpunkt der KI-Entwicklung stellen. Das Projekt startet während auf europäischer Ebene das Gesetz über Künstliche Intelligenz verhandelt wird, das einen Rechtsrahmen für "vertrauenswürdige" KI zu schaffen versucht, die "auf den Menschen ausgerichtet" ist. Dieses Gesetz wird verlangen, dass der Einfluss von KI-Systemen auf grundlegende Rechte und Werte geprüft und bewertet wird. Wie genau kann KI "auf den Menschen ausgerichtet" sein? Wie können wir sinnvoll beurteilen, ob und wie KI-Systeme grundlegende Rechte und Werte verletzen? Und wie können wir das Bewusstsein für diskriminierende Praktiken stärken und diese stoppen?

Carla Hustedt, Leiterin des Bereichs Digitalisierte Gesellschaft der Stiftung Mercator erklärt: "KI-Systeme sind soziotechnische Systeme. Ihre Gestaltung verlangt von uns, dass wir verschiedene Silos aufbrechen: Wir brauchen interdisziplinäre, internationale Forschung sowie die Zusammenarbeit von wissenschaftlichen Akteuren mit Akteuren aus Wirtschaft und Zivilgesellschaft. Genau hier setzt das Projekt an."

Das Projekt Wünschenswerte Digitalisierung ist in zwei Teile geteilt. Im ersten Teil untersuchen Forschende interkulturelle Perspektiven auf KI und grundlegende Rechte und Werte. So macht Dr. Cave deutlich: "Um die potenziellen Auswirkungen von

Algorithmen auf die Menschenwürde zu verstehen, müssen wir über den Code hinausblicken und Lehren aus Geschichte und Politikwissenschaft ziehen." In diesem Teil des Projekts werden Fragen aus zwei Blickwinkeln gestellt: aus anthropologischer Sicht (Wie wird unsere Vorstellung vom "Menschen" durch digitale Technologien beeinflusst und beeinflusst umgekehrt diese?) und aus intersektionaler Sicht (Wie prägen strukturelle und historische Ungerechtigkeiten die Entwicklung und den Einsatz von Technologie? Wie wirkt sich das auf Grundrechte und Werte aus?)

Die Teams in Cambridge und Bonn werden sowohl mit Kolleginnen und Kollegen aus Europa als auch mit Teams in Asien und Afrika zusammenarbeiten. So betont Prof. Gabriel: "Unabhängig von unseren spezifischen kulturellen Weltanschauungen stellen diese neuen Technologien unser Selbstverständnis als menschliche Wesen in Frage." Daher beschäftigt sich das Projekt mit grundlegenden anthropologischen Fragen über den Menschen im digitalen Zeitalter. Wie prägen unterschiedliche Ideen über den Menschen die Ansichten verschiedener Kulturen über wünschenswerte Digitalisierung?

Im zweiten Teil des Projekts Designing AI for Just and Sustainable Futures, werden Forschende beider Universitäten mit der KI-Industrie zusammenarbeiten, um Prinzipien für das Design und die Bildung von KI zu entwickeln, die Nachhaltigkeit und Gerechtigkeit in den Mittelpunkt des technologischen Fortschritts stellen. Prof. Aimee van Wynsberghe, Humboldt Professorin an der Universität Bonn, die das Bonner Team im zweiten Teil leiten und ihre Expertise zu Nachhaltigkeit einbringen wird, erklärt: "Nachhaltigkeit in all ihren Dimensionen - sozial, ökologisch, wirtschaftlich und technologisch - ist ein wesentlicher Wert bei der Entwicklung von KI. Nur wenn sie berücksichtigt wird, können diese Technologien unser Leben und unsere Welt verbessern."

Das Projekt startet im April 2022 mit einer Laufzeit von fünf Jahren; die erste der zweijährlichen Konferenzen findet Anfang 2023 statt. Das Kernteam, bestehend aus siebzehn Forschenden aus Cambridge und Bonn und Gastprofessorinnen und -professoren, wird mit einer Vielzahl nationaler und internationaler Partner eng zusammenarbeiten.