

# CLICK WORK

CLICK WORK  
CLICK WORK  
CLICK WORK



Für wenige Cent pro Aufgabe trainieren Millionen Menschen weltweit künstliche Intelligenzen

# CLICK WOORK CLICK WOORK DATA WOORK

Fotos sortieren, Texte klassifizieren: Der KI-Boom beruht auf der Leistung von »Klickarbeitern«. Die Wissenschaftlerin Milagros Miceli erforscht deren Arbeitsbedingungen

INTERVIEW: LARS-THORBEN NIGGEHOFF

**PA 8** Wenn die Öffentlichkeit von den Heerscharen von Arbeitern erfährt, die in armen Ländern für Technologiekonzerne arbeiten, ist oft die Rede von »Clickwork« oder »Micro-work«. Sie legen Wert auf den Begriff »Data Work«. Warum?

MILAGROS MICELI: Weil er am besten beschreibt, was diese Menschen machen. Ihre Arbeit besteht darin, Daten zu produzieren. Ihre Arbeit besteht kein KI-Modell auch nur ansatzweise funktionsfähig zu machen. Es ist kein Zufall, dass der aktuelle KI-Boom Plattformen »Mechanical Turk« im Jahr 2005 erfand, über die Unternehmen und Arbeiter verbunden werden. Nur dank solcher Modelle hat die KI-Industrie die nötigen Arbeitskräfte zur Datengenerierung, und das zu niedrigen Preisen. **Wo findet die KI-Industrie die Arbeitskräfte?** Data Worker gibt es grundsätzlich überall, auch in den USA und in Europa. Die wichtigsten Märkte finden sich aber in Entwicklungsländern, etwa in

154

Millionen

Menschen oder mehr arbeiten weltweit als Data Worker. Viele davon in Indien, Bangladesch, auf den Philippinen oder in Kenia

Indien und Bangladesch, aber auch auf den Philippinen und in Kenia. Das sind überwiegend ärmere Menschen. Weltweit arbeiten mindestens 154 Millionen Menschen als Data Worker. **Und wie genau produzieren diese Menschen Daten?**

Es gibt verschiedene Aufgabentypen, die sich aber auf vier Bereiche herunterbrechen lassen. Erstens: die Generierung neuer Daten. Dazu können die Data Worker das Internet scannen, also durchforsten, um bestehende Informationen zu sammeln und neu zu ordnen. Oder sie schießen Fotos, etwa von sich, von anderen Menschen und Gegenständen. Auch Sprachaufnahmen liefern sie an die KI-Unternehmen.

**Die dann damit ihre Algorithmen trainieren?**

Nein, erst mal müssen die Daten klassifiziert werden. Das ist die zweite Aufgabe, die die Data Worker übernehmen. Sie identifizieren zum Beispiel, was auf einem Bild zu sehen ist: Ist es eine Katze oder ein Hund? Dann versetzen sie

die Fotos mit entsprechenden Keywords. Sie bewerten auch, ob ein Bild zum Beispiel jugendfrei ist oder nicht, das ist wichtig, damit die Unternehmen wissen, ob sie es benutzen können. Damit können dann Algorithmen arbeiten. Aber meist sind sie ungenau. Die Feinjustierung übernehmen dann wieder die Data Worker, das ist Aufgabe drei. Sie prüfen etwa, wie präzise Empfehlungsalgorithmen sind, von Streaminganbietern zum Beispiel. So verhindern sie unpräzise Vorschläge. Wenn der Algorithmus nach einer Dokumentation über Affen zum Beispiel Filme mit schwarzen Menschen vorschlägt, ist das falsch und verteilend. Die Data Worker verhindern das mit ihrer Arbeit.

**Und der vierte Aufgabenbereich?**

Das sind Arbeiter, die KI-Systeme nachahmen, etwa Chatbots, wie viele Unternehmen sie im Kundenservice anbieten. Wir erwarten von Chatbots kein richtiges Gespräch, dadurch können Unternehmen ihren Kundenservice reduzieren und Geld sparen, selbst wenn noch ein echter Mensch mit den Kunden chattet. Außerdem ist KI gerade ein Hype-Thema, das Investoren anlockt. **Investoren und Kunden lassen sich aber nicht gern hinter das Licht führen.**

Nicht alle Firmen machen daraus ein Geheimnis. Sie framen es dann anders, sagen, dass die Data Worker den Algorithmus trainieren, bis der irgendwann selbstständig arbeiten kann. Ob das immer stimmt, ist eine andere Frage.

**Sie haben schon »Mechanical Turk« angesprochen. Sind solche Plattformen immer das Mittel der Wahl, wenn Firmen Data Worker suchen?**

Diese Plattformen sind die billigsten Anbieter, weswegen sie sehr erfolgreich sind. Sie funktionieren nach dem Gig-Economy-Prinzip: Die Arbeiter werden pro Aufgabe bezahlt, nicht pro gearbeitete Stunde. Das erhöht natürlich den Druck, möglichst schnell fertig zu werden, worüber sich die Plattformkunden wiederum freuen. Der Lohn bewegt sich meist bei wenigen Cent pro Aufgabe. Als Alternative gibt es noch Business-Process-Outsourcing-Anbieter (BPO). Das kann man sich wie eine moderne Variante eines Callcenters vorstellen. Die sind auch etwas teurer, weswegen Unternehmen auf diese vor allem zugehen, wenn sie höhere Qualität wünschen.

**Sie kritisieren vor allem die Kunden dieser Plattformen, aber auch die Anbieter immer**



**wieder in Ihrer Arbeit und in Interviews. Warum?**

Weil sie letztendlich Gewinn machen, indem sie die Arbeiter ausbeuten. Es gibt eine Reihe von Problemen. Gerade bei den Plattformen gibt es keine Kontinuität, die Data Worker hangeln sich von Aufgabe zu Aufgabe, haben keine Garantie, dass es morgen etwas zu tun gibt. Auch Transparenz fehlt. Die Arbeiter wissen nicht, wer der Kunde ist, welche Art von System sie trainieren sollen, und sie wissen nicht im Voraus, wie lange eine Aufgabe dauern wird. Entsprechend ist es schwer einzuschätzen, wie fair die Bezahlung ist. Auch Preisdiskriminierung ist Alltags in der Branche, nicht immer zahlen die Kunden für gleiche Arbeit gleichen Lohn, sondern zum Beispiel von Land zu Land einen unterschiedlichen. Aber auch die Aufgaben an sich sind manchmal problematisch.

**MILAGROS MICELI** ist Soziologin und Informatikerin. Sie leitet die Forschungsgruppe »Daten, algorithmische Systeme und Ethik« am Weizenbaum-Institut in Berlin. Sie erforscht vor allem ethische und soziale Auswirkungen von künstlicher Intelligenz, auch am Distributed AI Research Institute (DAIR), einer internationalen Organisation für konzernunabhängige KI-Forschung



»Gerade bei den Plattformen gibt es keine Kontinuität, die Data Worker hangeln sich von Aufgabe zu Aufgabe«

**2**  
US-Dollar pro Stunde erhielten Menschen in Kenia, die ChatGPT für die Firma OpenAI trainierten. Das ergaben journalistische Recherchen des US-Magazins »Time«

**Warum?**

Nehmen Sie zum Beispiel ethnische Kategorien aus den USA: Da gibt es etwa weiße Menschen, schwarze Menschen, Latinos und Menschen asiatischer Abstammung. Das sind Kategorien, mit denen ein Mensch in Bangladesch oft erst mal nichts anfangen kann. Wenn er dann für seine Arbeit zu lange braucht oder fehlerhaft arbeitet, wird er vom Plattform-Algorithmus zurückerufen und kriegt vielleicht bald keine Aufträge mehr. Abgesehen davon, dass es schon fragwürdig ist, von den Arbeitern zu verlangen, die Rasse von Personen anhand von Bildern zu erraten. Das ist bereits eine sehr umstrittene, aber leider auch häufige Aufgabe. Eine Möglichkeit, den Kunden zu fragen, haben sie in der Regel nicht.

**Werden die Data Worker auch mit Gewaltdarstellungen konfrontiert?**

Das kommt auch vor. Und das hat natürlich Auswirkungen auf ihre Psyche. Dann allerdings haben sie keine Unterstützung, keine Krankenversicherung, keine Möglichkeit, sich krankzuschreiben zu lassen. Also schauen sie sich für ein paar Cent pro Aufgabe weiter Fotos von Missbrauch oder Mord an.

**Man sollte meinen, dass die Anbieter ein Interesse daran haben, solche Zustände zu vermeinlichen. Wie bekommen Sie trotzdem Zugang für Ihre Forschung?**

In dieser Aufgabe werden Sie die **Rasse** der Personen auf den Fotos bestimmen. Sie sollen nur **eine** der folgenden Kategorien wählen:

- Weiß
- Afroamerikanisch
- Latinx oder hispanisch
- Asiatisch
- Indisch
- Unbeideutig

Das war einfach. Ich habe vor sechs Jahren angefangen, mich mit dem Thema zu befassen, und habe die entscheidenden Firmen angeschrieben. Sie waren sich nicht bewusst, wie problematisch ihre Beziehung zu den Datenarbeitern ist. Sie waren sogar geschmeichelt, dass jemand zu ihnen forsch. Als ich später meine Ergebnisse veröffentlichte, änderte sich das natürlich, das Verhältnis wurde schwieriger. Aber jetzt hatte ich direkten Zugang zu den Arbeitern und Arbeiterinnen. Es kamen auch zunehmend welche direkt auf mich zu, um mir von ihrer Situation zu berichten.

**Wie sieht Ihre Forschung konkret aus? Missbräuchliche Arbeitsbedingungen sind schwer in Daten zu fassen.**

Ich arbeite deswegen auch qualitativ, nicht quantitativ. Ich nutze also die individuellen Erfahrungen der Betroffenen, um Erkenntnisse zu gewinnen, anstatt statistische Erhebungen durchzuführen. Das bedeutet: viele Interviews führen, Verräte und Aufgabenbeschreibungen analysieren. Ich frage die Arbeiter, wie sie ihre Bezahlung vereinbaren und ob es ein Debriefing nach dem Job gab. Wie sah das aus? Mittlerweile arbeite ich auch mit Data Workern als Co-Researcher zusammen, da sie manche Dinge deutlich besser bewerten können als ich.

In dieser Aufgabe werden Sie **Hate Speech** identifizieren. Auf Basis des Textes müssen Sie entscheiden:

- **Hate Speech:** Ob der Nutzernamen Hassrede enthält.
- **None:** Ob KEINE Hassserfülle oder beleidigende Sprache in dem Gegebenen ist. (sic)

**DEFINITION:**

Wähle **Hate Speech**, wenn der Text etwas des Folgenden enthält:  
 • Diskriminierung, Verunglimpfung, Negativität oder Gewalt gegen eine Person oder eine Gruppe auf Basis des **geschützten Attributs**.  
 • Bezüge zu Hassgruppen oder Gruppen, die Menschen auf Basis eines geschützten Attributs attackieren.

Mitlagros Miceli trägt bei ihrer Arbeit viele Aufgaben zusammen, die Data Worker über die Plattformen von KI-Unternehmen vermittelt bekommen



## Die großen Vermittler

**CROWDSOURCING** Die wichtigste Plattform für die Vermittlung von Data Work ist »Amazon Mechanical Turk« (MTurk), das zur Web-Services-Division des US-Konzerns gehört. MTurk funktioniert nach einem ähnlichen Prinzip wie die Dienstleistungsunternehmen Uber oder Hepling: Auftraggeber und -nehmer können sich verknüpfen, Vertragskonditionen vereinbaren und die Bezahlung über die Plattform abwickeln. Der Name leitet sich vom »Schachtiraken« ab, einem vorgeblichen Schachroboter, den der Mechaniker Wolfgang von Kempelen im 18. Jahrhundert konstruierte – und hinter dem sich in Wahrheit ein menschlicher Schachgroßmeister verbarg. Die 2005 gestartete Plattform wird nicht nur von KI- und Technologieanbietern genutzt. Auch Forschende setzen auf die Plattform, um Teilnehmende für Studien zu gewinnen. Bei der Fähdung nach Vernisssten kam MTurk ebenfalls zum Einsatz, etwa bei der – letztlich erfolglosen – Suche nach dem Turing-Award-Gewinner Jim Gray. Der größte Konkurrent von MTurk ist die Plattform »Toloka«, die 2014 an den Start ging. Sie gehört zum russischen Technologiekonzern Yandex, bekannt für seine gleichnamige Suchmaschine. Der russische Begriff »toloka« stand in früheren Zeiten für eine Form gegenseitiger Unterstützung in Dorfgemeinschaften. Das Unternehmen veröffentlicht auch Datensätze und macht sie Forschenden zugänglich.

**Passiert das offiziell, oder müssen diese Personen fürchten, von ihrem Arbeitgeber belangt zu werden?**

Einige von ihnen fürchten tatsächlich Vergeltungsmaßnahmen und haben dennoch das Bedürfnis, über die Geschehnisse zu sprechen, und halten es für wichtig, dass die Öffentlichkeit von ihren Missständen erfährt. Deshalb ist es wichtig, sich als Forscher bewusst zu machen, was für die Arbeitnehmer auf dem Spiel steht, die mit uns sprechen und Forschungsprojekte mit uns durchführen. Wir schulden ihnen Respekt und Wertschätzung. Ohne die Datenbearbeiter, die ihren Arbeitsplatz aufs Spiel setzen, um mit mir zu sprechen, könnte ich meine Forschung nicht durchführen. In diesem Sinne gehört diese Forschung nicht nur mir, sondern auch ihnen.

**Wird Ihre Arbeit in der Öffentlichkeit wahrgenommen?**

Ja, mittlerweile. Aber das dauerte etwas. Seien wir ehrlich: Niemand liest wissenschaftliche Paper, die schreiben wir nur für unsere kleinen Kreise. Die Aufmerksamkeit kam, als ich begann, auf Veranstaltungen aufzutreten und Interviews zu geben. Dass parallel der große KI-Hype losbrach, hat natürlich geholfen.

**Wie reagierten die KI-Unternehmen wie OpenAI, Amazon oder Google auf Ihre Forschung?**

Ausgesprochen ablehnend. Sie reiten auf Details der Technologie herum, die ihrer Meinung nach nicht akkurat dargestellt sind, anstatt sich mit den eigentlichen Aussagen zu befassen. Auch Kollegen aus der Informatik haben mich nicht ernst genommen, weil ich eine Soziologin bin – und vermutlich auch, weil ich eine Frau bin.

**Wie sind Sie dem begegnet?**

Ich habe einen Doktor in Computerwissenschaften gemacht. Danach konnten sie mich nicht mehr ignorieren.

**Und die Plattformen?**

Die berufen sich nun auf Standards, die sie angeblich implementieren, oder schieben die Verantwortung auf die Dienstleister und Plattformen. Mittlerweile ist aber die Politik aufmerksam geworden, sowohl in Europa als auch in den USA. Eine Gruppe von Kongressabgeordneten hat auf Basis unserer Forschung den betroffenen Konzernen einen offiziellen Brief geschrieben und sie zu dem Thema befragt. Wir haben mit den Abgeordneten dafür auch direkt zusammengearbeitet. Das



»Ohne die Datenbearbeiter, die ihren Arbeitsplatz aufs Spiel setzen, um mit mir zu sprechen, könnte ich meine Forschung nicht durchführen«

ist aus zwei Gründen gut: Zum einen müssen die Firmen darauf antworten, also ein wenig Transparenz herstellen. Zum anderen können sie sich zukünftig nicht mehr darauf berufen, dass sie ja nichts gewusst hätten.

**Kann die Politik noch mehr bewirken?**

Ich will keine Ratschläge geben: ich will erreichen, dass die Politik den Arbeitern zuhört. Die Betroffenen wollen zum Beispiel nicht, dass die Firmen geschlossen werden. Sie wollen aber bessere Arbeitsbedingungen. Und die könnten Politiker durch Rechtsakte erzwingen. Es braucht eine Art Lieferkettengesetz für künstliche Intelligenz.

**Was für Verbesserungen wünschen sich die Arbeiter konkret?**

Zunächst mal fordern sie ein Ende der Preisdiskriminierung, gleiche Arbeit, gleicher Lohn. Sie wollen Eigentümerschaft für die Daten, die sie erzeugt haben, Bilder und Audiodateien zum Beispiel. Und sie wollen erfahren, was am Ende mit diesen Daten passiert. Und natürlich grundlegende Schutzmechanismen, Krankenversicherungen zum Beispiel.

**Sehen Sie Möglichkeiten für die Arbeiterinnen und Arbeiter, selbst aktiv zu werden?**

Absolut, sie haben Druckmittel. Diese Arbeit ist so grundlegend für Technologiekonzerne; wenn sie streiken würden, wäre das ein großes Problem. Und wir sehen erste Erfolge. In Kenia zum Beispiel haben sich Data Worker zu einer Gewerkschaft zusammengeschlossen, die sie nun auf ganz Afrika ausweiten wollen.

**Erwarten Sie, dass auch die Konzerne in Zukunft umdenken werden?**

Nicht unbedingt. Ich habe aber eine Hypothese, die, falls sie zutrifft, zu einem Umdenken führen könnte: Mein Eindruck ist, dass bessere Arbeitsbedingungen mit einer höheren Qualität der generierten Daten einhergehen. Das wiederum würde auch die Algorithmen verbessern. Meine ausführlichen Interviews haben den Eindruck bestätigt. Es wäre aber wichtig, sich dieser Frage auch quantitativ anzunehmen, also statistisch aussagekräftig. Dann wäre es im ureigenen Interesse der Unternehmen, die Arbeitsbedingungen zu verbessern. ■

**Impact Sourcing**

Einige Data-Work-Vermittler haben sich »Impact Sourcing« auf die Fahnen geschrieben. Die

Idee: Menschen, die in Entwicklungsländern am unteren Ende der ökonomischen Leiter stehen, sollen die Möglichkeit bekommen, einen existenzsichernden Lohn zu verdienen. Die Anbieter wollen also einen echten, positiven Unterschied für die Menschen machen, einen »Impact« erzielen. Ein großer Anbieter auf diesem Feld ist die US-amerikanische Sama Group. Allerdings geriet auch dieses Unternehmen in die Kritik: Es setze Data Worker unter unfairem und unsicheren Bedingungen ein, etwa als Moderatoren für soziale Medien

Stiefel, Sneaker oder Sandale? Eine KI-Objekte richtig erkennen kann, müssen Menschen sie wie in diesem Beispiel mit Ausdauer trainieren

What type of shoes do you see?

Boots  Sneakers  Other

What type of shoes do you see?

Boots  Sneakers  Other

What type of shoes do you see?

Boots  Sneakers  Other