

Bericht zum Tagungsthema der Fachgruppe Ethik und Soziale Arbeit: „Kann ein Algorithmus im Konflikt moralisch kalkulieren?“

„Kann ein Algorithmus im Konflikt moralisch kalkulieren?“ Diese Frage erkundet das Forschungsprojekt KAIMO am Beispiel der Kindeswohlgefährdung transdisziplinär mit den Disziplinen der Sozialen Arbeit, der Informatik und der Philosophie. Im Rahmen des Fachgruppentreffens Ethik und Soziale Arbeit diskutierten die Teilnehmenden mit den am Forschungsprojekt Beteiligten über das Thema der Urteilsbildung in moralischen Konfliktsituationen unter Hinzuziehung von künstlicher Intelligenz in der Sozialen Arbeit im Kontext der Kindeswohlgefährdung. Die Beurteilung und die Bearbeitung von Kindeswohlgefährdung ist dabei ein von der Sozialen Arbeit maßgeblich mitgestalteter Prozess im Rahmen der Schutzpflicht des Staates. Da gegebenenfalls als Ultima Ratio auch gegen den Willen in die Lebensführung interveniert wird, ist diese Situation immer als ethischer Ernstfall zu bewerten, der eine besondere Sorgfalt und Abwägung erfordert. Auftretende Zielkonflikte und das Handeln unter Unsicherheit stellen für die Fachkräfte Belastungen dar. Können also KI-Systeme in diesem Kontext die fachliche Arbeit verbessern?

Zunächst eröffnete Alexander Filipovic (Universität Wien) die Tagung mit einer Keynote zu den Themenfeldern Ethik, Digitalisierung und Politik. Dabei warf er die Frage auf, ob es legitim sei, künstliche Intelligenz als etwas darzustellen, dem man vertrauen könne. Vertrauen, so Filipovic richte sich doch auf Menschen. Dem widersprach Gunzelin Schmid Noerr: Menschen vertrauten durchaus auch Techniken, die sie nutzten, ohne diese zu verstehen. Auch wenn Filipovic dem zustimmte, warnte er doch vor der Übertragung von menschlichen Tätigkeiten wie lernen, urteilen und handeln auf die künstliche Intelligenz. Besser sei es, davon zu sprechen, wie KI-Systeme aufgestellt seien. Eine weitere Frage der Ethik im Kontext der Kultur der Digitalität sei, ob die Herausforderungen der Digitalität und der künstlichen Intelligenz mit den bisherigen ethischen Kriterien bewältigt werden können oder ob völlig neue entwickelt werden müssten. Dabei werde die Ethik von den einen für ein „Ethics Washing“ benutzt, um die technischen Innovationen zu legitimieren, von den anderen zum „Danger Reinforcing“, um die Technik abzulehnen. Von keinem der Lager wolle er sich vereinnahmen lassen.

Filipovic äußerte Skepsis gegenüber dem Standpunkt derer, die technische Innovationen vorantreiben wollen und vorbrächten, die Technik sei ja an und für sich neutral, es käme nur darauf an, wie man sie einsetze. Dem setzt Filipovic entgegen, dass Medien und Technik nicht nur das täten, was man ihnen auftrage, sondern dass sie die Aufgaben mitformulierten. Auch Tim Reiß (KHSB) wies daraufhin, dass die Form dem Inhalt nicht nur äußerlich sei, sondern neue Inhalte schaffe. Ein Beispiel dabei ist die Ersetzung der Mensch-Mensch Interaktion durch Mensch-Maschine Interaktion.

Es wurde festgehalten, dass im Entscheidungsprozesse Künstliche Intelligenz die makroskopische statistische Perspektive einbrächte. Das hermeneutische Fallverstehen kann dagegen nicht durch KI erbracht werden.

Für Filipovic steht fest, dass viele gesellschaftliche Bereiche vor einer Automatisierung stehen.

Anknüpfend an die Keynote führte Michael Reder (Hochschule für Philosophie München) in das vom Bayerischen Forschungsinstitut für digitale Transformation (BIDT) geförderte Forschungsprojekt KAIMO ein. Dabei will der Forschungsverbund sich der Frage zuwenden, wie sich soziale Prozesse durch den Einsatz von Technik verändern und wie künstliche Intelligenz gerade in hochemotionalen Situationen wie dem Kinderschutz einsetzbar sei. Das Forschungsvorhaben sei als ein iterierender und pragmatischer Prozess geplant, bei dem auch die Frage, ob ein KI gestütztes Assistenzsystem entwickelt werden könne, gestellt werde. Bei Fragen des Feldzugangs verwies Reder auf forschungsethische Fragen, z.B. inwieweit Forschende bei Hausbesuchen im Rahmen der Tätigkeit des Jugendamts dabei sein können.

Die Perspektive der Sozialen Arbeit

Jennifer Burghardt (TH Nürnberg) führt in das Themenfeld Soziale Arbeit und Digitalisierung im Kinderschutz ein und erörterte die Implikationen der digitalen Transformation von Kernprozessen der Sozialen Arbeit wie Diagnostik und Hilfekontaktmanagement. Die Digitalisierung allgemein diene bislang der Standardisierung und dem Controlling von Arbeitsprozessen. Die Implementierungsprozesse unterlagen dabei häufig der Doktrin: „Passt eure Prozesse unseren technischen Möglichkeiten“ an. Risiken bestehen in einer Deprofessionalisierung durch eingeschränkte Handlungs- und Gestaltungsspielräume und verzerrende Einflüsse auf die Urteilsbildung. Ebenfalls besteht die Gefahr, dass die Beziehungsgestaltung und die persönliche Dimension der Arbeit verloren gehen. Als Chancen nannte Burghardt eine evidenzbasierte Qualitätssteigerung und die Entwicklung neuer Angebote für die digitalen Lebenswelten. Die Einführung von **Onlineberatung** biete einen niedrigschwelligen Zugang samt ortsungebundenen Möglichkeiten der Kommunikation. Burghardt verwies auf Studien, die zeigten, dass auch textbasierte Beratung eine tiefgehende Auseinandersetzung mit emotionalen Themen ermögliche. Das Abfassen von Texten könne bereits eine therapeutische Wirkung entfalten. **Chatbots** böten die Möglichkeiten, Wissen bedarfsgerecht für die Nutzenden aufzubereiten und könnten auch die erreichen, die keine (Online)beratung in Anspruch nähmen. Im Zusammenhang mit schambehafteten Themen könnten sie auf eine größere Akzeptanz der Ratsuchenden treffen. Insbesondere im Bereich der **Risikodiagnostik** gewönnen die neuen Technologien an Bedeutung. Die Potenziale von mustererkennenden-statistischen Prognoseverfahren werfen die Frage auf, ob analog getroffene Prognosen tatsächlich besser seien. Berücksichtigt man den Handlungsdruck in der Kindeswohlgefährdung, so könnten Vorhersagen von Fachkräften mit KI-Unterstützung besser sein, als ohne. KI-Systeme könnten Informationen evidenzbasiert auswerten. Das Problem sei aber, dass Beratende keinen Zugriff auf die „Rohdaten“ hätten, als Beispiel wurden Fotos von Hausbesuchen genannt. Durch Beschreibungen könne es hier zu Verzerrungen kommen.

Vera Taube von der FHWS Würzburg benannte in ihrem Kommentar zum Impulsreferat als großes Risiko die Angst davor, aus Haftungsgründen gegen den Algorithmus zu votieren, selbst wenn die Fachkraft nicht überzeugt sei. Diese algorithmische Autorität könne zu einer automatisierten Entscheidung führen.

Mark Schrödter (Uni Kassel) merkte an, dass auch die Notwendigkeit von Prognosen legitimiert werden müsse. „Warum helfen wir nicht einfach allen?“, fragte er. Es müsse umso mehr prognostiziert werden, je weniger Hilfemöglichkeiten zur Verfügung stehen, denn dann müsste ja sehr genau unterschieden werden, wer jetzt Hilfe verdiene, bzw. wo interveniert werden müsse. Mark Schrödter verwies hier auf die unterschiedlichen Hilfesysteme in den USA und in Deutschland. Ein Wohlfahrtssystem, dem die Möglichkeit zur Verfügung steht, in unklaren Situationen Hilfeformen zuzuschalten, müsse nicht im gleichen Umfang auf Prognosen setzen. „Man kann auch erstmal eine sozialpädagogische Familienhilfe einsetzen und dann weitersehen...“

Die Perspektive der Informatik

Der Informatiker Maximilian Kraus (FHWS) verglich die Online-Tools im US-amerikanischen und europäischen Markt. Online-Tools in den Vereinigten Staaten suggerierten, man könne mit dem Einsatz der Tools jederzeit Kinderschutz als Service anbieten, quasi vom heimischen Sofa aus. Als zusätzliche Informationsquellen dienen hier teilweise Personendaten aus anderen Disziplinen wie zum Beispiel Strafverfolgung, Drogenaufsicht, Medizin oder Finanzwesen.

Onlinetools würden im europäischen Markt dagegen zurückhaltender beworben. Deutschsprachige Softwarelösungen im Bereich Kinderschutz fokussieren sich mehrheitlich auf die Prozessabbildung und die digitale Verwaltung von Fallakten und sind dabei nicht bis minimal mit amts-externen Datenquellen verknüpft. Während also in Ländern wie USA, Großbritannien oder Neuseeland mustererkennende-statistische Prognoseverfahren bereits erprobt werden, fokussieren sich die Angebote in Deutschland eher darauf, den Prozess abzubilden und Fallakten digital zu verwalten. Die im Projekt avisierten Assistenzsysteme dienen ebenfalls der systematischen Erhebung, Auswertung und Aufbereitung von Informationen und Daten zur Unterstützung der Entscheidungsprozesse von Fachkräften. Sie erheben nicht primär den Anspruch der abschließenden Feststellung einer (drohenden) Kindeswohlgefährdung (Ampelsystem), die eine entsprechende Intervention nach sich ziehen würde. Bei den Feldforschungen in deutschen Jugendämtern sei eine fragmentierte, wenig standardisierte Softwarelandschaft aufgefallen. Auch die fehlende Ausstattung mit Hardware (Diensthandys, Lap tops und Audioaufnahmegeräte) wurde festgestellt. Auf dieser Basis könnten nicht genügend standardisierte Trainingsdaten gewonnen werden.

Die Perspektive der Ethik

Rebecca Gutwald (Hochschule für Philosophie München) strich heraus, dass in den Entscheidungen, die im Kontext drohender Kindeswohlgefährdung gefällt werden müssen, wichtige Rechtsgüter wie Elternrecht und Kindeswohl abgewogen werden müssten. Neue Informationen können die Sachlage stark verändern. Die zu treffende Entscheidung sei deshalb nicht ideal und unsicher. In diesem Zusammenhang könnten KI-Systeme zwar rechnen, sortieren und nach Kriterien gewichten, sie könnten aber kein hermeneutisches Fallverstehen einbringen und auch nicht bewerten. Die Frage, ob ein Algorithmus moralisch kalkulieren könne, sei also mit Nein zu beantworten, sie könnten aber bei der Kalkulation hilfreich sein und bei der kritischen Reflexion der eigenen Urteilskraft unterstützen. Dabei ergäbe sich das Potenzial, dass KI-Systeme auf eigene blinde Flecken in der Urteilsbildung hinweisen.

Eine Gefahr bestehe, wenn KI-Systeme Entscheidungen nicht verbessern, sondern verschlimmern, wenn zum Beispiel Diskriminierung durch KI-Systeme eingeführt werden. Gutwald verwies auf die Studie von Eubanks (2018). Diese untersuchte das in den USA eingesetzte Tool zur Risikoanalyse „Allegheny Family Screening Tool“. Mit seinem Einsatz werde eigentlich Armut detektiert. Es kommt in diesem Zusammenhang zu einem Oversampling in den Daten. Familien, die andere Hilfen in Anspruch nehmen, werden stärker erfasst. Schlussendlich wird die Einschätzung produziert, dass schon die Tatsache, dass Hilfe gesucht wird, ein Risiko darstellt. Armut werde mit Vernachlässigung gleichgesetzt, eine faire und gleichberechtigte Beurteilung wird nicht mehr möglich, die Hilfeannahme stigmatisiert.

Auch andere gesellschaftliche Vorurteile wie rassistische oder sexistische Annahmen werden häufig von Algorithmen reproduziert und verstärkt, wenn die Datenbasis unreflektiert aus gegebenen Datenmengen übernommen wird. In KI-Systeme gehen also auch gesellschaftliche Vorurteile mit ein und werden durch die Verarbeitung im KI-System verschleiert. Es könne zu einer sogenannten „automation bias“ kommen, in dem Menschen der Maschine eine größere Autorität und Objektivität zuschrieben, weil sie als zuverlässiger und fehlerfreier gelten. Gutwald schließt: Das kann dazu führen, eigene kritische Gedanken oder ungewöhnliche Fälle, die nicht ins Schema passen, zu ignorieren und sich unkritisch auf die Maschine zu verlassen.

Rebecca Gutwald fragte, ob KI-Systeme nicht auch als Anti-Bias Hilfe eingesetzt werden könnten. So könnten KI-Systeme Fragen stellen, deren Beantwortung auf Verzerrung der Urteilskraft durch Vorurteile oder zu starke Komplexitätsreduzierung hinweisen können. In diesem Sinne könne man sich eine Chat-bot als Ergänzung der kollegialen Fallberatung denken oder als Unterstützung, wenn eine kollegiale Beratung nicht möglich ist.

Dabei muss darauf hingewiesen werden, dass die Transparenz der Entscheidung in einer demokratischen Gesellschaft für die betroffenen Bürgerinnen und Bürger unverzichtbar ist. Diese müssen wissen können, warum in einem Fall so entschieden worden ist. Sie sollten sich nicht blind fügen müssen. Entscheidungen können nicht in einer uneinsehbaren „Black box“ getroffen werden. In der Diskussion wurde dieser Gedanke aufgegriffen und gefragt, ob es nicht gegen die menschliche Würde verstoße, wenn man von einer Maschine beurteilt werde.

Stefanie Rosenmüller (FH Dortmund) regte an, verschiedene Urteilebenen zu differenzieren. Das wissenschaftliche Urteil von außen sei ein Urteil aus dem Abstand heraus, aus der reflektierenden Zuschauerperspektive. Die konkrete Urteilssituation brauche aber auch eine Klugheit aus der Akteursperspektive. Beide Perspektiven brauchen Verfahren, in denen jenseits des Typisierenden auch das Besondere an einer Situation herausarbeitet und partizipativ ein Urteil gebildet werden kann.

Theresia Wintergerst