




Programm „Voneinander Lernen“ auf dem Gebiet der Geschlechtergleichstellung

Künstliche Intelligenz und geschlechtsspezifische Verzerrungen in Einstellungs- und Auswahlverfahren Online-Seminar, 12.-13. November 2020

Zusammenfassung



*Die in der vorliegenden Veröffentlichung enthaltenen Informationen und Meinungen sind die der Verfasser*innen und geben nicht notwendigerweise die Auffassung der Europäischen Kommission wieder. Weder die Europäische Kommission noch Personen, die in ihrem Namen handeln, sind für die Verwendung der hierin enthaltenen Informationen verantwortlich.*



Diese Veröffentlichung wird unterstützt durch das EU-Programm Rechte, Gleichstellung und Unionsbürgerschaft 2014-2020.

Dieses Programm wird von der Europäischen Kommission umgesetzt. Sein Ziel ist es, einen Beitrag zur Weiterentwicklung eines Raums zu leisten, in dem die Gleichstellung und die Rechte von Personen – wie sie im Vertrag, in der Charta und in internationalen Menschenrechtsübereinkommen verankert sind – gefördert und geschützt werden.

Weitere Informationen unter: http://ec.europa.eu/justice/grants1/programmes-2014-2020/rec/index_en.htm

Einleitung

Im Rahmen des Programms „Voneinander Lernen“ auf dem Gebiet der Geschlechtergleichstellung (Europäische Kommission) fand am 12. und 13. November 2020 ein Online-Seminar zum Thema „Künstliche Intelligenz und geschlechtsspezifische Verzerrungen in Einstellungs- und Auswahlverfahren“ statt. Das Seminar wurde vom niederländischen Ministerium für Bildung, Kultur und Wissenschaft veranstaltet. Es sollte ursprünglich in Den Haag stattfinden und ein niederländisches innovatives Praxisbeispiel rund um einen „Hackathon“ präsentieren. Sowohl dieser als auch das Seminar selbst mussten aufgrund der laufenden Pandemiemaßnahmen ins Internet verlegt werden. Trotz dieser schwierigen Rahmenbedingungen waren Teilnehmer*innen aus 21 Mitgliedstaaten beteiligt: Dänemark, Deutschland, Estland, Finnland, Frankreich, Griechenland, Irland, Italien, Kroatien, Lettland, Litauen, Malta, die Niederlande, Österreich, Polen, Portugal, Rumänien, Schweden, Slowenien, Spanien und die Tschechische Republik.¹ Es handelte sich um Vertreter*innen von Regierungsbehörden, Universitätsfakultäten in den Bereichen Rechtswissenschaften, Gleichstellung und Informatik sowie von Forschungseinrichtungen und aus dem privaten Sektor. Ebenfalls vertreten waren die Europäische Kommission, das Europäische Institut für Gleichstellungsfragen und eine zivilgesellschaftliche Organisation namens „Women at the table“.

Das Seminar war eine hervorragende Gelegenheit, um die Herausforderungen und Chancen des Einsatzes von Künstlicher Intelligenz in Einstellungs- und Auswahlverfahren zu erörtern. Das Seminar konzentrierte sich auf zwei Hauptaspekte: 1) das potenzielle Risiko von (geschlechtsspezifischer) Diskriminierung durch Algorithmen bei der Personaleinstellung; und 2) die Sensibilisierung für das Problem geschlechterbezogener Voreingenommenheit im Zusammenhang mit Algorithmen. Die Europäische Kommission hatte einen koordinierten Ansatz angestrebt und Anfang 2020 ein Weißbuch² veröffentlicht. Darin wurde ein Rechtsrahmen für KI angeregt, um sicherzustellen, dass KI sicher, rechtmäßig und im Einklang mit den EU-Grundrechten zum Tragen kommt. Bei mangelhafter Transparenz und Robustheit von Algorithmen und maschinellem Lernen besteht die Gefahr, dass geschlechtsspezifische und andere Vorurteile übernommen und verstärkt werden. Im März 2020 hob der Beratende Ausschuss für Chancengleichheit von Frauen und Männern der Europäischen Kommission in einer Stellungnahme zu KI die Chancen und Herausforderungen für die

¹ Weitere Einzelheiten zu den am Seminar beteiligten Staaten können den Länderbeiträgen zum Seminar entnommen werden. Die Beiträge sind auf der Website des Programms verfügbar: https://ec.europa.eu/info/publications/artificial-intelligence-and-gender-biases-recruitment-and-selection-processes-online-seminar-12-13-november-2020_en

² https://ec.europa.eu/info/files/white-paper-artificial-intelligence-european-approach-excellence-and-trust_de

Geschlechtergleichstellung hervor.³ Die ebenfalls im März veröffentlichte Strategie für die Gleichstellung der Geschlechter 2020-2025⁴ bekräftigt das Potenzial von KI, erwähnt jedoch auch damit verknüpfte Herausforderungen in Bezug auf die Geschlechtergleichstellung und die Menschenrechte. Insofern knüpft das Seminar an die einschlägigen Maßnahmen und politischen Ziele der Europäischen Kommission an.

1. Hintergrund

Miriam Kullmann, thematische Expertin von der Wirtschaftsuniversität Wien, gab einen Überblick über die wichtigsten Arten von KI, die für die Personalbeschaffung zum Einsatz kommen. Sie skizzierte deren Vorteile aus Sicht der Unternehmen, insbesondere bei hohem Rekrutierungsvolumen. Daneben bergen solche Technologien jedoch auch vielfältige potentielle Ursachen von geschlechtsspezifischen Verzerrungen oder anderweitiger Diskriminierung. Deshalb ist es entscheidend, Werkzeuge zu entwickeln, um diese abzuwenden. Ein Algorithmus lässt sich definieren als „formal vorgegebene Abfolge von logischen Operationen, die dem Computer in Einzelschritte unterteilte Handlungsvorschriften für datenbezogenes Reagieren und damit für die Automatisierung von Entscheidungen bereitstellt“. Amazon steht beispielhaft für Unternehmen, die KI einsetzen, um die besten Anwärter*innen für Vorstellungsgespräche auszuwählen. Dabei kristallisierte sich allerdings heraus, dass der Algorithmus für das maschinelle Lernen männliche Kandidaten bevorzugte, weil er auf Lebensläufe aus Bewerbungen der letzten 10 Jahre trainiert worden war, die überwiegend von Männern stammten. Dieses Beispiel veranschaulicht sehr gut das Potenzial für Verzerrungen. Wenn die historischen Daten, die in ein algorithmisches System einfließen, im Hinblick auf die Bevölkerungsvielfalt – ob geschlechtsspezifische oder andere Merkmale wie ethnische Zugehörigkeit und Alter – Defizite aufweisen, besteht die Gefahr, dass entsprechende Stereotypen zu Handlungsvorschriften werden. Ohne valide Eingangsdaten kann das System auch keine aussagekräftigen Ausgangsdaten liefern.

Die thematische Expertin zeigte einige Wege auf, um der Gefahr geschlechterspezifischer Verzerrungen zu begegnen. Ein bedeutender Schritt ist insbesondere die Sensibilisierung auf staatlicher Ebene, aber auch unter Jurist*innen und IT-Fachleuten. Es braucht kreative rechtliche Lösungen um zu erwägen, wie vor allem EU-Nichtdiskriminierungsgesetze und die Datenschutz-Grundverordnung

³ https://ec.europa.eu/info/sites/info/files/aid_development_cooperation_fundamental_rights/opinion_artificial_intelligence_gender_equality_2020_en.pdf

⁴ <https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=COM%3A2020%3A152%3AFIN>

herangezogen werden könnten. Es gilt beispielsweise, mögliche Rechtsmittel ins Auge zu fassen, wenn jemand von Benachteiligung oder implizierter Diskriminierung durch KI betroffen ist. Dabei wäre auch zu überlegen, inwieweit ein Unternehmen für die Verwendung von Software haftbar gemacht werden kann, die eingebettete Vorurteile aufweist. Weitere Maßnahmen, die in Betracht gezogen werden können, sind Gefahrenbeurteilungen, Kontrollen oder ethische Vorgaben für die Bewertung KI-gestützter Ergebnisse.

Es folgte eine anregende Diskussion. Die Teilnehmer*innen bekräftigen den Stellenwert von Initiativen, die den Brückenschlag zwischen IKT (Informations- und Kommunikationstechnologie) und dem Themenkomplex Recht und Nichtdiskriminierung bzw. zwischen IT-Fachleuten und der allgemeinen Öffentlichkeit voranbringen wollen. Sie hielten fest, dass Frauen in der IKT und insbesondere in der KI nach wie vor unterrepräsentiert sind und dass es wichtig ist, die Geschlechtergleichstellung bereits in der Frühförderung zu thematisieren, um der Geschlechtersegregation auf dem Arbeitsmarkt entgegenzuwirken. Studierende und andere Arbeitssuchende benötigen Unterstützung bei der Vorbereitung und Präsentation von Bewerbungen.

Die Diskussion konzentrierte sich auf die Frage, wie die Qualität von Datensätzen verbessert werden kann, wenn diese selbst soziale Ungleichheiten widerspiegeln. Es wurde angemerkt, dass bei dem Versuch, die Datensätze zu diversifizieren, die Gefahr besteht, dass marginalisierte Gruppen angreifbarer werden, zum Beispiel wenn ethnische Zugehörigkeit einen bedeutenden potentiellen Diskriminierungsfaktor bildet. Damit verknüpft sind auch Fragen des Datenschutzes und der Überwachung. Die Teilnehmer*innen untersuchten die Frage, wie sich Daten diversifizieren und gleichzeitig die inhärenten Risiken für gefährdete Gruppen berücksichtigen lassen.

Die Hinterfragung der dem Modell zugrunde gelegten Daten wurde als ein weiterer möglicher Ansatz genannt. In diesem Zusammenhang wurde die kanadische Richtlinie aus dem Jahr 2019 zur Offenlegung des von der Regierung entwickelten Quellcodes als wertvolles Beispiel erwähnt. Aus strategischer Sicht ist es insofern möglicherweise sinnvoller, sich auf den öffentlichen statt auf den privaten Sektor zu konzentrieren.

Eine weitere Fragestellung lautete, ob infolge des Aufkommens von Algorithmen eine Anpassung der bisherigen rechtlichen Bestimmung von Diskriminierung angezeigt ist. So könnte in den Gesetzen zur Nichtdiskriminierung mehr Augenmerk auf vorbeugende oder wiederherstellende Funktionen gelegt werden. Dem wurde entgegengehalten, dass ein auf das Recht auf Rechtsbehelfe ausgerichteter Gesetzesrahmen meist ungeeignet ist, wenn technologische Weiterentwicklungen im Spiel sind. Der Rechtsbegriff der Nichtdiskriminierung ist weitgehend auf Einzelfälle zugeschnitten, während KI bereichsübergreifende und kollektive Situationen von Diskriminierung verursachen kann. Wenngleich sich mittelbare Diskriminierung auf ganze Kollektive beziehen kann, unterliegt sie Regelungen, die im Kontext von KI ebenfalls kaum greifen. Ein Beispiel ist die Beweislast. In diesem Zusammenhang

erörterten die Teilnehmer*innen, ob es angebracht wäre, den Unternehmen das „Risiko“ zu übertragen, über die Ergebnisse automatisierter Entscheidungsabläufe aufzuklären, und welche spezifischen Strategien dafür geeignet wären. Möglicherweise wäre es sinnvoll, entweder einen gesonderten Regelsatz für Nichtdiskriminierung auszuarbeiten, um spezifische KI-bezogene Herausforderungen zu behandeln, oder aber die bestehende Antidiskriminierungsrichtlinie abzuändern bzw. um eigene KI-Bestimmungen zu ergänzen.

2. Niederländisches Praxisbeispiel: „Hackathon“ über KI und geschlechtsspezifische Verzerrungen in der Personalbeschaffung

2.1 Ziel und Ablauf des „Hackathons“

Das niederländische Ministerium für Bildung, Kultur und Wissenschaft und die ihm unterstellte Direktion für die Gleichstellung von Frauen, Männern und LGBTI-Personen haben strategische und operative Ziele für die Stellung der Frau auf dem Arbeitsmarkt bis 2025 festgelegt. Angestrebt wird, die finanzielle Unabhängigkeit von Frauen (Verdienst von mindestens 100% des Mindestlohns) zu gewährleisten, da Frauen in den Niederlanden im Vergleich zu anderen EU-Mitgliedstaaten eine anhaltend hohe Teilzeitarbeitsrate aufweisen. Ein weiteres Ziel besteht darin, in Unternehmensvorständen mehr Diversität herzustellen.

Die niederländische Regierung ist besorgt, dass neue Technologien potentiell auch dazu beitragen, Ungleichheiten zwischen den Geschlechtern auf dem Arbeitsmarkt zu verfestigen. Mit noch laufenden Gesetzesentwürfen soll dafür gesorgt werden, dass Arbeitgeber*innen beim Einsatz neuer Technologien gewährleisten müssen, dass keine Diskriminierungen hervorgerufen werden. Das Ministerium hat Studien zu KI und geschlechtsspezifischen Arbeitsmarktaspekten in Auftrag gegeben, die 2021 abgeschlossen werden sollen. Es liefert Beiträge zu verschiedenen politischen Dokumenten, ethischen Rahmenwerken und Gestaltungsprinzipien. Es hat eine koordinierende Funktion inne und führt zu diesem Themenkomplex einen ständigen Dialog mit anderen Ministerien sowie dem Privatsektor.

Das Niederländische Institut für Menschenrechte veröffentlichte im Oktober 2020 einen bahnbrechenden Bericht (auf Niederländisch) über KI und Diskriminierung in Auswahl- und Einstellungsverfahren. Dabei wurde die potenziell diskriminierende Beschaffenheit der ursprünglichen Konzeption von Algorithmen hervorgehoben. Der Bericht bewog das Ministerium, trotz des Gesundheitsnotstands an den Plänen für den „Hackathon“ festzuhalten.

Der als brückenbauende und Veranstaltung für den Wissensaustausch angelegte „Hackathon“⁵ fand von 30. Oktober bis 1. November online statt. In einem Wettbewerb sollten Lösungen erarbeitet werden, um Verzerrungen bei der Erfassung und Kennzeichnung automatisierter Systemprozesse für die Beurteilung und Auswahl von Jobkandidat*innen zu verhindern. Dabei wurden fünf Risikofaktoren zu Tage gefördert, die zur Entstehung von Verzerrungen beitragen können: a) die Definition von „guter“ Leistung oder Eignung; b) die Erfassung und Verarbeitung von Ausbildungsdaten; c) die Auswahl von Variablen; d) Näherungsgrößen; und e) Ausblendung. Die 50 Teilnehmenden waren auf zehn interdisziplinäre Teams aufgeteilt, darunter Data Scientists, Fachleute für Personalbeschaffung, Business Analysts, Unternehmer*innen, Jurist*innen, Sozialwissenschaftler*innen sowie Beauftragte für Diversität und Integration. Am Ende des „Hackathons“ sollten die Teams ihre Vorschläge präsentieren. Eine Jury wählte die besten Beiträge aus.

2.1.1 Diskussion

Es wurde die Frage untersucht, wie sich Diskriminierung im Einzelfall erkennen lässt, ohne den Schutz vertraulicher Daten preiszugeben. Algorithmen sind überaus komplex, funktionieren wie eine „Blackbox“ und entwickeln sich zudem ständig weiter. Für die Erkennung von Vorurteilen ist es deshalb möglicherweise einfacher, bei einer Analyse des Outputs – also der Ergebnisse – anzusetzen. Es wurde zudem festgehalten, dass das „Blackbox“-Syndrom in der realen Welt existiert und es schwierig ist, geschlechtsspezifische Diskriminierung selbst dann auszumachen, wenn keine Algorithmen im Spiel sind.

Die Diskussion kam auch auf die Bedeutung der rechtlichen Abgrenzung von Diskriminierung zurück. Schutz und Rechtsmittel gegen Diskriminierung setzen voraus, dass Informationen über das Auswahlverfahren zugänglich sind. Wenn undurchsichtige „Blackbox“-Systeme zum Tragen kommen, haben Opfer von Diskriminierung keine wirksamen Rechtsmittel – ein Widerspruch zu den Grundprinzipien des EU- sowie des einzelstaatlichen Rechts. In Finnland findet keine Koordinierung der Kontrollinstanzen für Gleichstellung und Datenschutz statt. Obwohl die Möglichkeit bestünde, Folgenabschätzungen durchzuführen, hat dieses Themenfeld bislang keine Beachtung gefunden. In den Niederlanden stellt die Aufsichtsbehörde für soziale Angelegenheiten Überlegungen dazu an, welche Überwachungsfunktion im Rahmen der vorgeschlagenen neuen Gesetzgebung in Frage kommt. Darin sollen Unternehmen verpflichtet werden, beim Einsatz von KI für Diskriminierungsschutz zu sorgen.

Es wurde auf die Komplexität einer umfassenden Erkennung geschlechtsspezifischer Diskriminierung hingewiesen. Das geht so weit, dass Frauen sich mitunter eine Selbstzensur auferlegen. Es wurde ins Treffen geführt, dass sich Diskriminierung bei

⁵ <https://www.youtube.com/watch?v=3D1oH5gV-80&feature=youtu.be>

der Verwendung von KI im Allgemeinen mit *ungleichen Auswirkungen* (ein US-Konzept) oder *indirekter Diskriminierung* begründen ließe. In dieser Hinsicht könnten die UN-Leitprinzipien für Wirtschaft und Menschenrechte als nützlicher Anknüpfungspunkt dienen: Sie verschieben die Risikoperspektive – und rücken das „Risiko für den Menschen“ anstelle des „Risikos für das Unternehmen“ in den Fokus. Es wurde ergänzt, dass *ungleiche Auswirkungen* vorliegen, wenn eine scheinbar neutrale Politik oder Praxis Einzelpersonen aufgrund geschützter Merkmale in unangemessener Weise benachteiligt. Folglich könnte die Bewertung unterschiedlicher Auswirkungen in das Gesetz oder zumindest in die Folgenabschätzungen zu KI aufgenommen werden. Eine Möglichkeit wäre die gesetzliche Verpflichtung, dass diese Abschätzungen von der zuständigen, mit der Durchsetzung von Antidiskriminierungsbestimmungen beauftragten Aufsichtsbehörde durchzuführen ist.

Die Teilnehmer*innen überlegten, welche Art von Politik oder Instrumenten erforderlich ist, damit Unternehmen sich der Risiken bewusst werden oder dazu angehalten werden können, Maßnahmen zu ergreifen. Die neuseeländische „Algorithmus-Charta“ wurde als wertvolles Beispiel für politische Maßnahmen angeführt.

Kontrollmechanismen könnten sich ebenfalls als nützlicher Weg erweisen, um die Art, Quantität und Qualität von Daten sowie deren Trainingsprozess nachvollziehbar zu machen. Das würde auch Unternehmen zugutekommen, die den Nachweis erbringen möchten, dass ihre Systeme nicht voreingenommen sind. In Schweden stellte die Ombudsperson für Diskriminierung bei der Untersuchung eines konkreten Falls fest, dass das Unternehmen die Verantwortung trägt, falls sich die verwendeten KI-Systeme als diskriminierend erweisen. Das hat bewirkt, dass einige Unternehmen ein stärkeres Problembewusstsein entwickeln und geeignete Maßnahmen ergreifen.

Schließlich wurde auf Beispiele von Unternehmen verwiesen, die innovative Pilotprojekte angestoßen haben. In Österreich schließt ein Unternehmen in seinem Algorithmus das Geschlecht als Kriterium aus. Die Chancengleichheit im Auswahlverfahren habe sich dadurch verbessert. Die schwedische Innovationsagentur Vinnova versucht, KI so umzugestalten, dass sie aktiv zur Förderung der Geschlechtergleichstellung beiträgt. Das niederländische Institut für angewandte Technologie wiederum wickelt ein Pilotprojekt ab, bei dem fiktive voreingenommene und nicht voreingenommene Daten herangezogen werden, um die Ergebnisse von Einstellungs- und Auswahlverfahren zu vergleichen.

2.2 Prämierte Lösungen des „Hackathons“

Das Programm des „Hackathons“ fußte auf dem „Design-Thinking“-Konzept und der menschenzentrierten Innovation. Das Programm umfasste Mentoring-Sitzungen und runde Tische mit einer Vielzahl von Sachverständigen. Jedes Team arbeitete einen Vorschlag aus, der von der Jury nach vier Erfolgskriterien bewertet wurde:

a) Mehrwert und Wirkung; b) Kreativität und Innovation; c) Machbarkeit und d) Präsentation.

Der dritte Preis (Publikumspreis) ging an „Simplify 360“. Es handelt sich um ein Instrument, das Führungskräfte kleiner und mittlerer Unternehmen (KMU) dabei unterstützen soll, auf Basis einer Persönlichkeitsbeurteilung integrative und vielfältige Teams zusammenzustellen. Ein Kernelement ist ein Kennzahlen-Dashboard für inklusives Matching. Der zweite Preis ging an „Balancing the Scale“. Diese Lösung unterstützt den Aufbau eines vielfältigen Bewerbungspools, indem sie sicherstellt, dass im Screening-Prozess ausschließlich entscheidungsrelevante und nachvollziehbare Daten verwendet werden, basierend auf Erfahrung und der Messung von Wissen und nicht auf personenbezogenen Merkmalen wie Familienstand, Staatsangehörigkeit und Alter. Den ersten Preis erhielt „FAIR“. Die Idee ist die Entwicklung eines transparenten, modularen Bewertungsinstruments, bei dem Personalverantwortliche für jede Stelle fünf Qualifikationen vorgeben, anhand derer Bewerber*innen dann eine Selbsteinschätzung vornehmen können. Die KI ermittelt unter Wahrung der Geschlechterdiversität die 20 besten Kandidat*innen, aus denen die Personalverantwortlichen jene Gruppe auswählen, die zu einem Vorstellungsgespräch eingeladen wird.

Die beiden prämierten Teams erhielten ein Preisgeld sowie die Möglichkeit, ihre Ideen im Rahmen eines von der niederländischen Regierung geförderten Inkubationsprogramms weiterzuentwickeln. Zu den Lehren aus dem „Hackathon“ zählt, dass es notwendig ist, den Teilnehmenden ausreichend Hintergrundinformationen zur Verfügung zu stellen, um ein gemeinsames Verständnis für die Komplexität der Aufgabenstellungen zu gewährleisten. Ein wichtiges Erfordernis ist auch, den Teams weiterführende Möglichkeiten zum Ausbau ihrer Vorschläge zu bieten, um aus den Ergebnissen wirklich Nutzen zu ziehen.

2.2.1 Diskussion

Die Teilnehmer*innen bekundeten großes Interesse an dem „Hackathon“-Experiment und diskutierten dessen Stärken und Schwächen. Sie meldeten teilweise Bedenken an, was den Einsatz von Selbsteinschätzungen oder anderen Leistungsbewertungen und psychologischen Tests anbelangt. Diese seien nicht geschlechtsneutral, sondern wiesen tendenziell eine einseitige Ausrichtung auf männliche Persönlichkeiten und Männlichkeiten auf. Ein weiterer Einwand richtete sich gegen den Vorschlag, personenbezogene Merkmale gänzlich zu eliminieren. Das käme einer „übermäßigen Objektivität“ gleich. Es wurde auch angemerkt, dass einige der Vorschläge Raum für Subjektivität böten und zu willkürlichen Ergebnissen führen könnten. Der KMU-Fokus berge das Risiko eines zu kleinen Anwärter*innen-Pools. Es wurde darauf hingewiesen, dass sich das „Mehrwert“-Kriterium in der Beurteilung durch die Jury zweideutig auslegen lässt: Aus der „Unternehmens- und Markt“-Perspektive oder aus der Warte der „Gerechtigkeit“. Im zweiten Fall wäre der Zweck der AI dem Gleichheitsgrundsatz untergeordnet, was voraussetzen würde, dass

Chancengleichheit im Hinblick auf das unternehmerische Interesse stärker ins Gewicht fällt als Effizienz.

Die Teilnehmer*innen diskutierten den Umgang mit dem Mutterschaftsurlaub und die Frage, wie ein allgemeiner algorithmischer Erfahrungsparameter aussehen könnte, um einen Ausgleich zwischen verschiedenen Berufserfahrungen vorzunehmen. Sie nahmen mit Interesse zur Kenntnis, dass eines der Teams sich diesem Thema gewidmet hatte.

Es wurde die Frage der Unternehmenskultur angesprochen, die von den Führungskräften tendenziell auch dann angewendet wird, wenn sie diskriminierende Praktiken enthält. In allen präsentierten Fällen geht das Einstellungsverfahren mit menschlicher Intervention einher, was seiner behaupteten Objektivität zuwiderlaufen könne. Dem wurde entgegengehalten, dass die Personalbeschaffung grundsätzlich – und sei es nur aus ethischen Belangen – mit menschlicher Beteiligung erfolgen solle. Die Teilnehmer*innen überlegten, ob Algorithmen zum Zug kommen könnten, um diese Unternehmenskultur zu neutralisieren, oder ob es sinnvoller wäre, sich auf einen langfristigen Kulturwandel und die Stärkung des Bewusstseins für Sozialverantwortung zu konzentrieren. Im Zusammenhang mit den Streitkräften wurde auf die Herausforderung hingewiesen, bei der Personalbeschaffung nicht nur mehr Diversität zu erreichen, sondern die Beschäftigten angesichts der impliziten oder expliziten Kultur dann auch zu halten.

In einem weiteren Diskussionsfeld stand die Frage im Mittelpunkt, was unter einem „fairen“ Algorithmus zu verstehen ist und wie sich ethische Begrifflichkeiten in Algorithmen umsetzen lassen, um neutrale Kriterien zu verstärken. Die Komplexität und Bedeutung „entzerrender“ Algorithmen wurde angesprochen. Dabei gebe es vielgestaltige Modelle und eine Vielzahl mathematischer Definitionen von Gerechtigkeit. Insofern als „Fairness“ ein vielschichtiger und subjektiver Begriff ist, wägen die Teilnehmer*innen ab, ob sich aus EU-Sicht eine einheitliche Grundlage abstecken ließe oder ob es besser wäre, die Menschenrechte als Basis heranzuziehen. In diesem Zusammenhang wurde auch festgehalten, dass Algorithmen in Genusssprachen ein höheres Verzerrungsrisiko aufweisen können. Um Transparenz und Nachvollziehbarkeit zu untermauern, wurde vorgeschlagen, dass Datensätze mit einem Informationsblatt einhergehen. Darin könnten – ähnlich wie beim Verkauf von Elektronikgeräten – ihr Zweck, ihre Zusammensetzung, der Erhebungsprozess und die empfohlenen Einsatzbereiche aufgeschlüsselt werden.

Mehr Aufschlüsse über die Verwendung von impliziten Kriterien und Näherungsgrößen wurden als wichtig erachtet, da sie häufig zu Verzerrungen führen können, so dass etwa Personen im Fortpflanzungsalter oder aufgrund ihrer sexuellen Orientierung ausgeschlossen werden. In diesem Zusammenhang wurde eingeworfen, dass Postleitzahlen in Großstädten als Näherungsgröße dienen können, um Bewerber*innen mit Migrationshintergrund auszuschließen.

Schließlich wurden Zertifizierungssysteme als nützliche Methode erwähnt, um bewährte Verfahrensbeispiele auszuzeichnen. Viele öffentliche und private

Online-Seminar, 12.-13. November 2020

Unternehmen sind bestrebt, Auswahlverfahren fair, angemessen und verhältnismäßig zu gestalten. Sie stehen dem zunehmenden Einsatz von KI mit großer Vorsicht gegenüber. Ein Zertifizierungsprozess würde Unternehmen dabei helfen, sich auf einheitliche Instrumente zu einigen und den Nachweis über die Belastbarkeit ihrer Auswahlssysteme zu erbringen.

2.3 Ausblick: Nachwirkung des „Hackathons“ und Weichenstellungen für die Zukunft

Der „Hackathon“ war nicht dazu gedacht, Lösungen zu entwickeln, sondern ein Baustein in den Bemühungen, alle Beteiligten einzubeziehen und das Problembewusstsein zu verstärken. Der Bericht über den „Hackathon“ wird weitflächig verbreitet werden. Es wird damit gerechnet, dass die großen Unternehmen die Ergebnisse mit Interesse aufnehmen und zum Anlass nehmen, sich an der Entwicklung neuer Prototypen zu beteiligen.

Mit den beiden Siegerteams wird es weitere Gespräche darüber geben, in welcher Form die Inkubatorprogramme mit Unterstützung des Innenministeriums ausgebaut werden können. Die anderen teilnehmenden Teams werden ermutigt, ihre Vorschläge weiterzuentwickeln; dabei werden ihnen Fachleute für Datenmonitoring beratend zur Seite stehen.

Der „Hackathon“ war Teil eines breiter angelegten Programms, das eine Debatte über die Rolle der KI in der Gesellschaft anregen soll. Im Rahmen dieser Sensibilisierungsinitiative ist eine Reihe von Informationsmaterialien wie z. B. Merkblätter geplant.

3. Zentrale Diskussionspunkte des Seminars

Zahlreiche Teilnehmer*innen merkten zwar an, dass der Einsatz von KI in Einstellungs- und Auswahlverfahren offenbar noch nicht weit verbreitet ist, gingen aber davon aus, dass sich dies schon in naher Zukunft ändern dürfte. Die übereinstimmende Feststellung tendierte dahin, dass das **potenzielle Risiko von geschlechtsspezifischer Diskriminierung bei KI-gestützten Einstellungsverfahren** auf verschiedenen Ebenen angegangen werden muss. Dabei gilt es, nationale Gegebenheiten und Unterschiede in Bezug auf Diskriminierung zu berücksichtigen. Die Teilnehmer*innen sprachen sich für einen interdisziplinären Ansatz aus, wie er im Online-Seminar und beim „Hackathon“ zum Tragen kam. Es herrschte auch Einigkeit darüber, dass einheitliche, auf die Menschenrechte abgestellte Definitionen zu Aspekten wie Voreingenommenheit und Diskriminierung begrüßenswert wären.

Es gab eine breite Diskussion über die Einsatzmöglichkeiten von KI, ihre potenziellen Vorteile und ihre Auswirkungen auf die Zukunft der Arbeit. Es wurde als wichtig erachtet, eine angemessene Balance zwischen Regulierung und Kontrolle einerseits und der Förderung von Forschung und Innovation andererseits zu finden.

Online-Seminar, 12.-13. November 2020

Rein technisch wurde Kontrolle als eine praktikable Option befunden, obwohl hinsichtlich der Transparenzerfordernisse und der Offenlegung von Quellcodes Einschränkungen festzuhalten seien, die zum Teil mit dem Geschäftsgeheimnis und dem Schutz geistigen Eigentums zusammenhängen. Kontrollen könnten von externen Unternehmen oder von bestehenden staatlichen Aufsichtsorganen durchgeführt werden. Die Herausforderung bestünde jedoch darin, die am wenigsten sichtbaren Fälle von Diskriminierung ans Tageslicht zu fördern. Ein weiteres Erfordernis ist ggf. die Schulung der Kontrollorgane, zumal es sich um ein komplexes Themenfeld handelt. Die Zusammenarbeit zwischen den staatlichen Gleichstellungs- und Datenschutzstellen könnte gefördert werden. Die Teilnehmer*innen zogen noch weitere Ansätze in Betracht, etwa die Fokussierung auf die Einschätzung der Fairness von Ergebnissen oder den Rückgriff auf Gütesiegel für Unternehmen.

Es gab Diskussionen über die rechtlichen Aspekte von Kontrollmechanismen und darüber, ob die notwendigen juristischen Instrumente auf nationaler oder EU-Ebene existieren bzw. wie sie ggf. geschaffen werden könnten. Einige Teilnehmer*innen vertraten die Ansicht, dass Diskriminierung im Zusammenhang mit KI derzeit aus gesetzlicher Sicht nicht angemessen berücksichtigt ist. Andere merkten an, dass Gesetzesentwürfe zur Regulierung des Einsatzes von KI vorliegen, wobei es seitens der IT-Unternehmen erheblichen Widerstand gegen derartige Maßnahmen gebe. Es wurde erörtert, welche Vorteile freiwillige Vorgaben für Unternehmen bzw. öffentliche Intervention auf nationaler oder Unternehmensebene mit sich bringen.

Eine Feststellung lautete, dass Gewerkschaften hinsichtlich des Zugangs zu Informationen über Personalbeurteilungen und Personalbeschaffung eine wichtige Funktion zukommt. Es sei angezeigt, Gewerkschaftsvertreter*innen und Betriebsratsmitglieder über die Rolle und mögliche Gefahren von KI zu schulen. Kollektivverhandlungen, möglicherweise in bestimmten leitenden Schlüsselsektoren, könnten sich als sehr wertvoll erweisen.

Sensibilisierung bezüglich der geschlechtsspezifischen Voreingenommenheit von Algorithmen wurde als wesentliches Erfordernis angesehen, da ein allgemeines Informationsdefizit herrsche und viele Menschen KI und Codes für „neutral“ hielten. Insofern besteht ein Bedarf an Aufklärung in der breiten Öffentlichkeit und bei allen Interessengruppen. Es gilt außerdem sicherzustellen, dass bestehende Unternehmensrichtlinien zu Diversität und Inklusion aktualisiert und verbreitet werden. In diesem Zusammenhang wurde betont, wie wichtig es sei, eine integrative Unternehmenskultur zu schaffen, in der Meinungsverschiedenheiten oder unterschiedliche Stimmen gehört und respektiert werden können.

Schulungen in verschiedenen Bereichen wurden als wesentlich erachtet, unter anderem für Personalmanager*innen und IKT-Teams, die am Einstellungsverfahren beteiligt sind, ebenso wie für Gewerkschaftsmitglieder, Jurist*innen und Gleichstellungsfachleute sowie für die breite Öffentlichkeit. Ein Austausch bewährter Verfahrensbeispiele wäre sehr hilfreich.

Abschließend wurde festgehalten, dass Diskriminierung im Zusammenhang mit KI an die in der realen Welt bestehenden Probleme anschließt. Wie die Teilnehmer*innen betonten, gilt es die Bemühungen zur Förderung der Geschlechtergleichstellung im gesamten Bildungssystem zu verstärken, beginnend bei der frühkindlichen Förderung. Ein zentrales Anliegen müsse sein, mehr junge Mädchen zu einer Laufbahn in MINT- und IKT-Berufen zu ermutigen und junge Frauen, die diesen Weg einschlagen, mit zielgerichteten Mitteln zu unterstützen.

4. Schlussfolgerungen und Empfehlungen

Die Vielschichtigkeit der Problemstellung, die sozialwissenschaftliche, rechtliche und Fragen des Unternehmensmanagements gleichermaßen berührt, lässt eine breit gefasste Methodik – nach dem Vorbild eines Geschäftsplans – als zweckmäßigen Ansatz für das weitere Vorgehen erscheinen.

Die Teilnehmer*innen sprachen sich für folgende Maßnahmen im Rahmen eines interdisziplinären, menschenrechtsbasierten Konzepts aus:

- Ergänzende Studien, um das Ausmaß des Einsatzes von KI in Einstellungs- und Auswahlverfahren deutlicher zu erfassen.
- Ausgabe einer Checkliste oder eines Leitfadens zur Unterstützung des öffentlichen Sektors sowie privater Unternehmen, die den Einsatz von KI für die Personalbeschaffung in Erwägung ziehen. Führungskräfte sollen ein besseres Verständnis für maßgebliche Aspekte erhalten, die in Bezug auf Diskriminierung aufgrund des Geschlechts oder sonstiger Merkmale zu beachten sind.
- Aufbau einer Datenbank oder eines Archivs mit bewährten Verfahrensweisen und Pilotprojekten, die von Unternehmen und anderen Akteur*innen durchgeführt wurden.
- Ausarbeitung von Strategien u. a. mit Schwerpunkt auf das Kontrollwesen sowie die Notwendigkeit, rechtliche Rahmenbedingungen auf die neuen technologischen Realitäten zu übertragen.
- Aufsichtsorgane in den Bereichen Datenschutz und Gleichstellung auf nationaler und EU-Ebene sollen ermutigt werden, in Fragen zu Vorurteilen in KI-gestützten Einstellungs- und Auswahlssystemen zusammenzuarbeiten.
- Schulungen für Führungskräfte, Personalverantwortliche und Gewerkschaftsvertreter*innen, um das Bewusstsein für mögliche Verzerrungen in KI-gestützten Personalbeschaffungssystemen zu verstärken. Empfehlenswert wären zudem Schulungsangebote zu Diversität und Rechtsfragen für technische Fachkräfte.
- Maßnahmen zur Förderung der Inklusivität, damit mehr Frauen IKT-Berufe ergreifen.