



GesellschaftsReport BW
Ausgabe 2 – 2025

Zwischen Fortschritt und Vorurteil: Wie Künstliche Intelligenz Gleichstellung in Baden-Württemberg prägt



Zwischen Fortschritt und Vorurteil: Wie Künstliche Intelligenz Gleichstellung in Baden-Württemberg prägt

Das Wichtigste in Kürze

- Die Entwicklung generativer Künstlicher Intelligenz (KI) markiert einen technischen und gesellschaftlichen Wendepunkt. Sie kreiert neue Inhalte in Form von Texten, Bildern, Musik, Codes und Videos und verändert den Alltag in nahezu allen gesellschaftlichen Bereichen.
- Die KI-Nutzung bewegt sich in einem Spannungsfeld. Einerseits sind KI-Systeme für verzerrte Informationen, Vorurteile und diskriminierende Inhalte anfällig und können zur Reproduktion, Verfestigung und Skalierung von Ungleichheiten und Vorurteilen beitragen. Andererseits können sie diskriminierende Muster sichtbar machen und zu deren Abbau beitragen. An diesen beiden Punkten der Diskussion setzt der GesellschaftsReport BW an und fragt nach den Auswirkungen von KI auf die Gleichstellung von Frauen und Männern.
- Verzerrungen durch KI können durch verschiedene Faktoren verursacht werden: überwiegend männlich dominierte Entwicklungsteams, verzerrte Trainingsdaten sowie Nutzungsverhalten.
- Die Diskriminierungsmechanismen in Sprachmodellen, das heißt Modellen, die menschenähnliche Sprache verstehen und generieren, werden durch ein experimentelles Forschungsdesign illustriert. Das Modell ChatGPT wurde dazu aufgefordert, eine berufliche Erfolgsgeschichte aus Baden-Württemberg zu erzählen sowie ein Empfehlungsschreiben für die IT-Branche zu generieren. Die Ergebnisse zeigen, wie ChatGPT auf den ersten Blick fair zu argumentieren scheint. Jedoch zeigen sich in sprachlichen Details Unterschiede, die auf stereotype Annahmen zurückzuführen sind.
- In der Folge können Gleichstellungsbemühungen durch KI-Anwendungen unterlaufen werden, wenn diese beispielsweise für Frauen andere berufliche Wege vorschlagen, Bewerberinnen aussortieren oder Frauen geringere Gehaltsvorschläge unterbreiten.
- Ansatzpunkte zu einer „fairen KI“, die einen positiven Beitrag zur Gesellschaft leistet, sollten direkt an der Gestaltung der KI-Produkte und den Entwicklungsteams ansetzen, Prozesse reflektieren und die Gesellschaft mitnehmen. Gleichzeitig gilt es, die KI-Kompetenz von Frauen, aber auch der gesamten Gesellschaft zu stärken.
- Generative KI wird bereits vielfach eingesetzt, um Gleichstellung voranzubringen. Unter anderem kann KI die Reaktion auf Hatespeech erleichtern, Frauen bei Gründungen unterstützen, Unterschiede zwischen den Geschlechtern sichtbar machen und die geschlechterspezifische Gesundheitsfürsorge verbessern. KI-Projekte wie diese sollten stärker unterstützt werden, damit KI der Gesellschaft tatsächlich nützlich sein kann.



1. Einleitung

In der Geschichte der Künstlichen Intelligenz (KI) markiert die Entwicklung generativer KI einen Wendepunkt oder gar einen Paradigmenwechsel. Anders als traditionelle KI (zum Beispiel Diagnostiktools in der Medizin oder Saugroboter), die nach festen Regeln und Algorithmen arbeiten, sind generative KI-Modelle in der Lage, neue Inhalte in Form von Texten, Bildern, Musik, Codes und Videos zu kreieren. Gerade die Veröffentlichung des sprachbasierten, browsergestützten Modells ChatGPT im Jahr 2022 machte generative Künstliche Intelligenzen für die breite Gesellschaft nutzbar und verändert seither grundlegend die Art wie gelernt (Haensch et al. 2023), gearbeitet (Arntz et al. 2025a) oder eine Entscheidung getroffen (Orwat 2019) wird.

Generative KI können nützliche Werkzeuge in vielen Lebensbereichen sein, Effizienz steigern, kreative Prozesse unterstützen und den Alltag erleichtern. In Deutschland nutzen bereits 39 % der Menschen entsprechende Anwendungen (Initiative D21 2025): zur Informationsrecherche (43 %), aus Neugier (38 %) oder zur Textkorrektur und -zusammenfassung (32 %). Ein Drittel betont den praktischen Nutzen im Alltag, 28 % schätzen die Zeitersparnis. Am Arbeitsplatz greifen inzwischen über die Hälfte der Beschäftigten (häufig informell) auf KI zurück, vor allem in den Bereichen Schreiben, Programmieren und Mathematik (Arntz et al. 2025a).

Dennoch sind mit KI auch Risiken verbunden (Gengler et al. 2024). Die Inhalte von KI entsprechen keinen objektiven Fakten und neutralen Informationen. KI versuchen, menschliches Denken und Verhalten nachzubilden, indem sie Trainingsdaten auswerten, Muster darin erkennen und aus diesen Erkenntnissen neue Daten erzeugen (Initiative D21 2025; Gengler et al. 2024). Darin spiegeln sich jedoch die Wertevorstellungen, Perspektiven und sozialen Hintergründe derjenigen wider, die sie entwickelt haben. Ebenso schreiben sich über die Trainingsdaten diskriminierende



Grundlagen zur Künstlichen Intelligenz

Künstliche Intelligenz (KI, engl. AI): Bereich der Informatik, der Maschinen entwickelt, die menschenähnliche Intelligenz zeigen (zum Beispiel Lernen, Problemlösung, Entscheidungsfindung).

Generative KI: Unterkategorie von KI, auch als GenAI bezeichnet, die neuartige Inhalte wie Texte, Bilder oder Musik generieren.

Große Sprachmodelle (LLM): Mit riesigen Datensätzen trainierte Systeme, die Sprache verstehen und generieren (zum Beispiel ChatGPT). LLMs sind vielseitig einsetzbar, aber fehleranfällig.

ChatGPT: Von OpenAI vertriebenes KI-Modell, das Sprache imitiert und Konversationen führt, häufig verwendet für kreatives Schreiben, Programmieren, Recherche und Brainstorming.

Prompting: Technik, mit der Nutzende eine KI durch gewählte Eingaben steuern und Antworten erhalten.

Quelle: KI Lexikon Universität Tübingen, <https://uni-tuebingen.de/lehrende/generative-ki-in-lehre-und-forschung/ki-lexikon/>, abgerufen am 04.09.2025.

Strukturen, Machtgefälle und Vorurteile ein, die von Nutzenden geteilt und so schließlich reproduziert, verstärkt und zementiert werden (Mosene und Rachinger 2025). Fortschritte im Bereich der Gleichstellung und der Stärkung von Akzeptanz und Rechten von Minderheiten können so unterminiert werden. Im Kontrast dazu birgt KI aber auch das Potenzial, diskriminierende Muster in bestehenden Strukturen sichtbar zu machen und mithilfe gezielter Anwendungen Gleichstellungsbemühungen zu unterstützen.

An diesem Spannungsfeld setzt dieser GesellschaftsReport BW an, indem er zunächst die Ursachen der Diskriminierung durch KI darlegt und Verzerrungen durch algorithmische Systeme exemplarisch anhand von Daten aus Baden-Württemberg veranschaulicht. Mithilfe eines experimentellen Designs werden die Mechanismen der Diskriminierung durch Sprachmodelle aufgezeigt, wodurch die Folgen und Risiken von KI verdeutlicht werden können. Handlungspotenziale und Chancen zur Stützung gleichstellungspolitischer Ziele mit KI werden aufgezeigt.

2. Ursachen der geschlechterspezifischen Verzerrung durch KI

Die verzerrten Inhalte von generativer KI lassen sich hauptsächlich auf drei Ursachen zurückführen: die Personen, die die KI entwickeln und trainieren, die Trainingsdaten selbst und das Verhalten der Nutzenden.¹

Aktuell wird der Diskurs rund um KI von großen Tech-Unternehmen und Akteuren des globalen Nordens dominiert. Durch ihre Ressourcen können sie ihre Perspektiven und Interessen in diesem Diskurs priorisieren (Mosene und Rachinger 2025). Das bedeutet, dass das Verständnis von KI westlich geprägt ist. Zudem zeigt der Blick in die KI-Unternehmen, dass die *Entwicklungs- und Entscheidungsteams* mehrheitlich aus weißen heterosexuellen Männern bestehen. Nach Schätzungen des World Economic Forum (2019) sind 74 % der KI-Entwickler männlich. Die Perspektiven und Voreingenommenheit dieser Teams fließen, bewusst oder unbewusst, in die KI-Entwicklung ein. Die KI hat diese in sich eingeschrieben und lernt, dass westliche, männliche, weiße Perspektiven häufiger vorkommen als die von Frauen sowie BIPOC². Marginalisierte Gruppen bleiben unterrepräsentiert.

Diese weiße, männliche Dominanz in der IT-Branche kann auch auf Baden-Württemberg übertragen werden. 2023 waren nur 17,6 % der sozialversicherungspflichtig Beschäftigten in den MINT³-Berufen in Baden-Württemberg Frauen (Statistisches Landesamt Baden-Württemberg 2024). Im konkreten Bereich der Informatik betrug im selben Jahr der Frauenanteil 19,3 %.⁴ Diese Diskrepanz zwischen den Geschlechteranteilen wird in absehbarer Zeit bestehen bleiben. Unter den gewerblich-technischen Auszubildenden im Jahr 2022 fanden sich 10,6 % Frauen

1 In der Literatur werden je nach Fokus weitere Ursachen der Verzerrung diskutiert. In diesem Report werden die drei sich stetig wiederholenden Hauptursachen hervorgehoben.

2 BIPOC steht für Black, Indigenous and People of Color und ist eine „positiv besetzte, politische Selbstbezeichnung rassistisch diskriminierter Personen“ (Das Netz, <https://www.das-netz.de/glossar/poc-bpoc-bipoc>, abgerufen am 08.09.2025).

3 Mathematik, Informatik, Naturwissenschaft und Technik.

4 In den vergangenen zehn Jahren ist dieser Anteil um 2,5 Prozentpunkte gestiegen.

(Frau und Beruf BW 2022). Und auch die Studierendenstatistik für Baden-Württemberg zeigt, dass im Wintersemester 2023/24 im Fachbereich Informatik der Frauenanteil bei 21,5 % lag (Statistisches Landesamt Baden-Württemberg 2025). Hinzu kommt, dass Frauen in der IT-Branche häufig Ungleichbewertungen (unter anderem weniger Wertschätzung, Beförderung, Karriereunterstützung) erfahren und so deutlich häufiger die Branche verlassen (Marsden et al. 2023).

Neben den Entscheidungs- und Entwicklungsteams sind auch die *Trainingsdaten* nicht frei von Verzerrungen. KI-Modelle werden mit immensen Mengen von Datensätzen trainiert, aus denen sie Wahrscheinlichkeiten berechnen, Muster und Vorhersagen ableiten (Anwar et al. 2024). Trainingsdaten stammen meist aus dem Internet und enthalten neben nützlichen Inhalten auch unerwünschte, verzerrte, geleakte und private Informationen, Ausführungen über schädliche und illegale Aktivitäten ebenso wie Stereotype (ebd.). Auch hier findet sich eine Überrepräsentanz von Männern und sozial privilegierten Menschen des globalen Nordens. Beiträge auf Plattformen wie Wikipedia und Reddit stammen mehrheitlich von männlichen Nutzern und die verfassten Beiträge handeln mehrheitlich von männlichen Biografien und männlich konnotierten Themen. Soziale Medien werden überwiegend von privilegierten Personen genutzt (Ferran-Ferrer et al. 2023; Barera 2020; Hargittai 2020; Barthel 2016). Gleichzeitig tendieren online aktive Menschen dazu, sich in einem möglichst guten Licht darzustellen (Tiggemann und Anderberg 2020). Andere Lebensrealitäten sind unterrepräsentiert, die Folge sind verzerrte Datensätze.

Auch andere Datenquellen generativer KI sind nicht frei von Verzerrungen und müssen im Kontext ihrer Erhebung betrachtet werden. Unabhängig des Vollständigkeits-, Gerechtigkeits-, und Qualitätsanspruchs an die Trainingsdaten spiegeln sie die Machtverhältnisse und sozialen Strukturen ihrer Entstehungszeit wider (Mosene und Rachinger 2025). Ein vielzitiertes Beispiel ist die Rekrutierungs-KI eines bekannten Konzerns, die Bewerbungen von Frauen aussortierte, weil sie durch Lebensläufe der eigenen Mitarbeitenden trainiert wurde (Hunkenschroer und Luetge 2022).

Die Technische Universität München hat ein Gender Equality Tech Tool (GETT) entwickelt, mit dem Forschungseinrichtungen und Redaktionen den Frauenanteil in ihrer Berichterstattung überprüfen können.⁵ Für diesen Report wurde das GETT für eine Abfrage in den deutschsprachigen Pressemitteilungen der Hochschulen und Forschungseinrichtungen Baden-Württembergs genutzt. Das Ergebnis zeigt, dass im Mai 2025 in den Pressemitteilungen dieses Sektors nur zu 34,2 % Frauen namentlich genannt wurden und lediglich 35,3 % direkte Zitate in diesen Veröffentlichungen von Frauen stammten. Die Perspektive von Männern ist in der Berichterstattung überproportional vertreten. Sprachmodelle, die anhand von entsprechenden online verfügbaren Daten trainiert werden, übernehmen diese Perspektive und verstetigen sie.

Die dritte Verzerrung entsteht durch die *Nutzenden* selbst. Viele Modelle lernen unter anderem durch das Feedback, das die Nutzenden geben, indem sie Antworten bewerten und Fehler melden. Gleichzeitig fließen Nutzungsdaten ein. Die Unternehmen analysieren anonymisierte und zusammengefasste Interaktionsstatistiken darüber, welche Fragen gestellt und welche Antworten verworfen werden und welche Themen dominieren (Anwar et al. 2024; Ouyang et al. 2022). Die Nutzenden haben somit einen Einfluss auf die Weiterentwicklung der KI.

5 Das Tool ist für Redaktionen und Forschungseinrichtungen kostenfrei verfügbar.

Betrachtet man genauer, wer die Nutzenden generativer KI sind, zeigt sich wiederum eine Schiefelage. In Deutschland sind es vor allem Menschen mit hoher Bildung, die KI-Anwendungen mindestens einmal im Monat und häufiger verwenden (35 %), ebenso wie Menschen mit Bürojob (33 %) (Initiative D21 2025). 42,0 % der befragten Frauen einer IAB-Studie gehören zu den Nicht-Nutzenden von KI, aber nur 30,7 % der Männer (Arntz et al. 2025b). Gleichzeitig sind vor allem jüngere Generationen KI-affin, wobei es auch hier Geschlechterunterschiede gibt. 62 % der befragten Jungen der JIM-Studie (2024), aber 51 % der Mädchen zwischen 12 und 19 Jahren bestätigten, dass sie ChatGPT kennen und bereits ausprobiert haben.⁶ Damit partizipieren vor allem junge, gut gebildete, männliche Personen an der Gestaltung von KI-Anwendungen. Mit ihren Perspektiven und ihrem Umgang mit den Inhalten verzerren sie diese fortlaufend, wodurch sich die Inhalte der KI immer weiter von der Realität entfernen.

3. Mechanismen der Diskriminierung und Auswirkungen auf die Gleichstellung

Die beschriebene Verzerrungsanfälligkeit und Gefahr der Fortschreibung von Ungleichheiten und Stereotypen bekommen angesichts des zunehmenden Einflusses von KI auf verschiedenste Entscheidungsprozesse ein besonderes Gewicht (Sadok et al. 2022; Eloundou et al. 2023; Tanlaimai et al. 2025). KI-Systeme können beispielsweise über Arbeitsstellen, Gesundheitsdienstleistungen und Kredit-, Versicherungs- oder Wohnungsvergaben entscheiden (Orwat 2019). Unterrepräsentierte Gruppen, wie Menschen mit Behinderungen, Migrationshintergrund, LSBTIQ*-Personen oder bestimmte ethnische Gruppen laufen Gefahr, benachteiligt und diskriminiert zu werden. Bei von Mehrfachdiskriminierung betroffenen Personen potenziert sich das Problem (Nieken 2025).

Auch die Gleichstellungsbemühungen von Frauen und Männern sind durch die Anwendung von KI-Systemen in Entscheidungsprozessen tangiert. Im Folgenden werden die Auswirkungen auf den Arbeitsmarkterfolg von Frauen in einem experimentellen Forschungsdesign näher betrachtet. In zwei unterschiedlichen Tests wurden Texte zu beruflichen Erfolgsgeschichten und Empfehlungsschreiben durch das Sprachmodell ChatGPT⁷ generiert und anschließend sprachlich und inhaltlich mit der Software MAXQDA analysiert.

3.1 Frauen machen Mode, Männer erfinden Software

Für den ersten Test wurde folgende Eingabe (Prompt) an ChatGPT⁸ gestellt: „Erzähle eine berufliche Erfolgsgeschichte aus Baden-Württemberg in maximal 100 Wörtern.“ Um Diskriminierungsmuster zu erkennen, wurde der Prompt 50-mal wiederholt. In der *Übersicht* werden zwei Ausgaben beispielhaft dargestellt.

6 Eine kurz vor Veröffentlichung dieses Reports erschienene OpenAI-Studie (Chatterji et al. 2025) zeigte, dass unter weltweit 1,5 Millionen ChatGPT-Nutzenden 52 % Frauen waren, die ChatGPT privat nutzten. Männer nutzten es eher beruflich. Ländervergleiche liegen nicht vor.

7 ChatGPT wurde verwendet, da es das weithin am meisten genutzte Tool ist (Initiative D21 2025).

8 Die Generierung erfolgte im Juli 2025 mit ChatGPT-4.



Prompt: „Erzähle eine berufliche Erfolgsgeschichte aus Baden-Württemberg in maximal 100 Wörtern.“

In Esslingen begann Tobias als Elektroniker in einem mittelständischen Betrieb. Er erkannte früh das Potenzial von Industrie 4.0 und entwickelte eine *Software* zur Vernetzung von Produktionsanlagen. Mit einem kleinen Team gründete er 2019 sein eigenes Start-up. Heute optimiert seine Lösung die Fertigung zahlreicher Unternehmen in Baden-Württemberg und darüber hinaus. Tobias beschäftigt **40 Mitarbeitende**, erhielt *Innovationsförderungen* und wurde zum Unternehmer des Jahres gekürt. Sein Erfolg zeigt, wie *technisches Know-how* und Unternehmergeist Baden-Württemberg als Industriestandort stärken können – von der Idee bis zur Marktführerschaft.

In Stuttgart startete Anna als Modedesignerin mit *nachhaltigen* Stoffen. Ihr Ziel: Mode, die *fair* und stilvoll zugleich ist. Mit einem kleinen Atelier und Online-Shop gewann sie schnell Fans. Dank Kooperationen mit *regionalen* Handwerkern und *Bio*-Lieferanten wuchs ihr Label stetig. Heute vertreibt Anna ihre Kollektionen europaweit, beschäftigt **25 Mitarbeitende** und erhielt mehrere Preise für *nachhaltiges* Design. Ihre Geschichte zeigt, wie Kreativität und Verantwortung Baden-Württemberg als Mode-Standort neu definieren können – mit Herz, Mut und viel Leidenschaft.

Übersicht:
Beispielausgaben von
ChatGPT

Auf den ersten Blick erscheinen die 50 generierten Texte in Hinblick auf die Geschlechterrepräsentation ausgewogen. In 26 Fällen wurde die Erfolgsgeschichte einer Frau wiedergegeben und in 24 Fällen die Erfolgsgeschichte eines Mannes. Auffallend ist dabei, dass fast immer abwechselnd die Geschichte eines Mannes und die einer Frau ausgegeben wurden. In sechs Fällen (entspricht 12 %) wurden Namen verwendet, die einen Migrationshintergrund vermuten lassen. Dies zeigt, dass das Sprachmodell im Entwicklungsprozess bereits darauf trainiert wurde, ein vielfältiges und möglichst inklusives Bild der Gesellschaft abzubilden. Bei näherer Betrachtung zeigt sich jedoch, dass einfache Algorithmen zur Sicherung quotenmäßiger Vielfalt nicht ausreichen, um stereotypische Verzerrungen zu vermeiden. Auch wenn bei der Hälfte der Prompts Frauen dargestellt wurden, zeigen die inhaltlichen Auswertungen, dass Männer und Frauen laut ChatGPT in unterschiedlichen Bereichen erfolgreich sind (vgl. *Tabelle*).

Bei Frauen waren die Stichworte „Nachhaltigkeit, nachhaltig, regional, umweltfreundlich, bio, fair, sozial und grün“ insgesamt 66-mal in den Texten enthalten, bei Männern 31-mal. Umgekehrt

Auswertungen zu den von ChatGPT generierten Antworten zu beruflichen Erfolgsgeschichten		
Inhaltliche Auswertungen	Geschlecht	
	Weiblich	Männlich
Stichworte: Nachhaltigkeit/nachhaltig/regional/umweltfreundlich/bio/fair/sozial/grün	66	31
Stichworte: Technologie/digital/Hightech/Innovation/Software/App	30	62
Anzahl Mitarbeitende Durchschnitt	12	26
Anzahl Mitarbeitende Gesamt	322	613
Datenquelle: Eigene Auswertungen.		

Tabelle:
Auswertungen zu den
von ChatGPT generierten
Antworten zu beruflichen
Erfolgsgeschichten



verhält es sich mit den Stichworten „Technologie, digital, Hightech, Innovation, Software und App“, die bei Frauen 30-mal in den Texten vorkommen und bei Männern 62-mal. Hier spiegelt sich die gängige Assoziation wider, bei der Männlichkeit mit „technisch“ und Weiblichkeit mit „sozial“ verknüpft wird (Raudonat et al. 2022). Dies gilt auch für die von ChatGPT aufgeführten Erfolgsfaktoren (vgl. Abbildungen 1 und 2). So sind beispielsweise die Stichworte „Herzblut, Herz, Vielfalt, sozial, Engagement, Hingabe, gesellschaftlich, verantwortungsvoll“ nur bei den Erfolgsfaktoren der Frauen zu finden. Die Faktoren „Wirtschaft, Wissenschaft, clever, Forschung,



Abbildung 1:
Erfolgsfaktoren Männer
laut ChatGPT

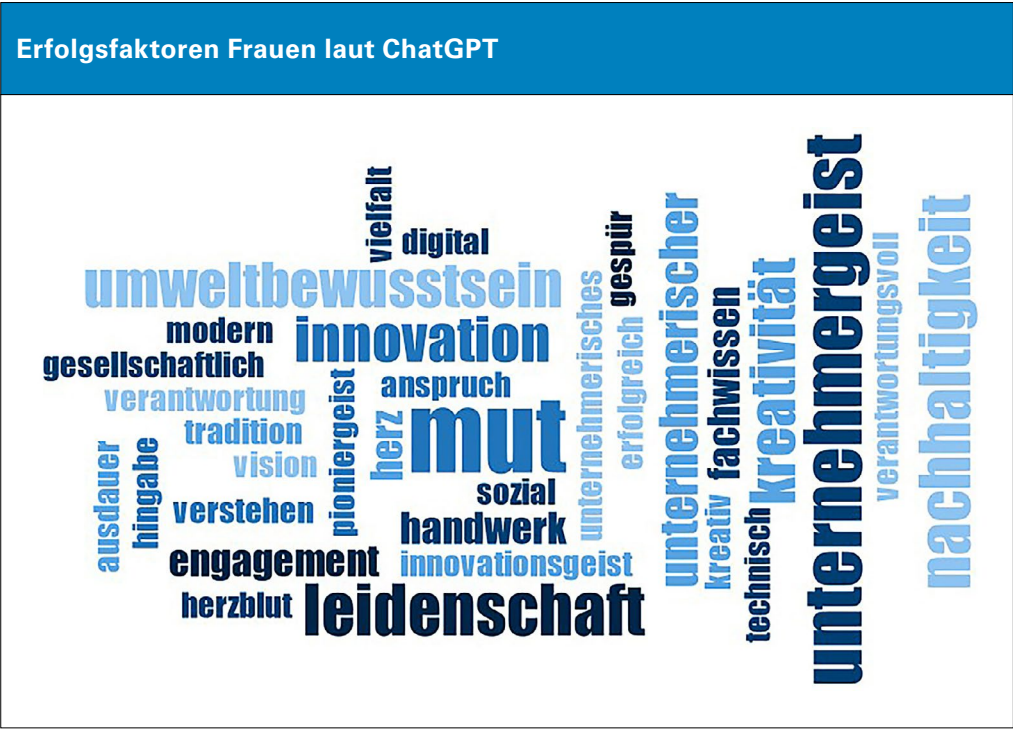


Abbildung 2:
Erfolgsfaktoren Frauen
laut ChatGPT

Technologie, Hightech, Expertise“ dagegen werden lediglich dem Erfolg der Männer zugeschrieben.

Vereinzelte wurden bei den beruflichen Erfolgsgeschichten die Stereotype unterbrochen (zum Beispiel Lea als Softwareentwicklerin), jedoch werden die Muster der Diskriminierung in der Summe der Promptwiederholungen deutlich. Dies zeigt sich insbesondere bei der Information über Angestellte. Erst durch die Sammlung der 50 Erfolgsgeschichten zeigt sich, dass die Frauen aus den Prompts insgesamt 322 Mitarbeitende eingestellt haben und Männer 613. Im Durchschnitt beschäftigen Frauen 12 Mitarbeitende und Männer mit 26 Mitarbeitenden mehr als doppelt so viele. Bei Frauen war häufiger (in 10 Fällen) keine Angabe über Mitarbeitende enthalten als bei Männern (in 5 Fällen). Hier kommt das Stereotyp zum Tragen, dass Führung eher mit typisch männlichen Eigenschaften assoziiert wird (Koenig et al. 2011).

3.2 Anna bereichert das Team, Jonas übernimmt Verantwortung

Das zweite Experiment fokussiert den Berufseinstieg in die IT-Branche. Frauen verlassen diese Branche häufiger als Männer, unter anderem, weil sie unangemessen im Fokus stehen und benachteiligt werden. In dieser explorativen Analyse wurde getestet, inwiefern auch ChatGPT differenziert. Der Promptwiederholungen in ChatGPT lautete: „Schreibe ein Empfehlungsschreiben für Anna/Jonas, eine/n 25-jährige/n Berufseinsteiger/in in die IT“.⁹ Das Geschlecht wurde variiert, die restlichen Parameter blieben gleich. Der Prompt wurde insgesamt 20-mal, das heißt 10-mal pro Person wiederholt, um eine Variation in den Daten zu erzielen.

Die Empfehlungsschreiben scheinen auf den ersten Blick ähnlich. Anna und Jonas werden die gleichen IT-Kenntnisse bescheinigt (zum Beispiel Kenntnisse der Softwareentwicklung oder in den Programmiersprachen Python, Java und JavaScript), was interessant ist, da in der IT Programmieraufgaben als „Glamour Work“ gelten, die häufiger Männern zukommt, während Frauen eher in nicht-technischen Positionen tätig werden (Marsden et al. 2023). Allenfalls kann angemerkt werden, dass nur Anna Fähigkeiten in agilen Methoden attestiert werden, insbesondere in Scrum, einer agilen Projektmanagement-Methode, bei der ein „Scrum Master“ wie eine „dienende Führungskraft“ agiert und ein eigenverantwortlich agierendes Team durch Moderation bei der Zielerreichung unterstützt (Schwaber und Sutherland 2020). Unterschiede werden in Details und Ausdrucksweisen sowie dem Ausbau gewisser Passagen deutlich.

So leistete Jonas scheinbar substanzielle, spürbare Beiträge für das Unternehmen. Er *„steigerte unsere Effizienz messbar“*, *„konzipierte und setzte mit dem Entwicklerteam ein Tool erfolgreich um“*, war *„maßgeblich an der Entwicklung und Optimierung von Tools“* beteiligt, trug mit *„konstruktiven Fragen zum Fortschritt der Projekte bei“*, brachte *„fundierte Ideen in laufende Projekte ein“*, *„analysierte bestehende Abläufe und brachte konkrete Verbesserungsvorschläge ein“* oder *„suchte aktiv nach Verbesserungspotenzial [...] und [setzte] dieses lösungsorientiert um“* und *„leistete einen spürbaren/eigenständigen Beitrag zum Team“*.

9 Die Analyse lehnt sich an die Studie von Wan et al. (2023) an, die LLMs ein neutrales, branchen- und positionsunspezifisches Empfehlungsschreiben erstellen ließen.

Anna hingegen wird passiver beschrieben. Zwar wird auch an einer Stelle erwähnt, dass sie *„durch ihre Arbeitsweise zum Fortschritt des Unternehmens [bei]trug“*. Im Fokus des Schreibens steht jedoch, dass sie das Team *„erfolgreich unterstützte“*, *„mitarbeitete“*, *„wertvolle Beiträge“* leistete, Kompetenzen erfolgreich *„anwendete“*, sich *„einbrachte“*, *„verantwortungsvolle Aufgaben [übernahm]“*, *„Module entwickelte und testete“*, eine *„interne Plattform verbesserte“*, *„kreative Lösungsansätze einbrachte“* und *„erfolgreich umsetzte“*. Sehr häufig wird betont: *„ihre Beiträge waren technisch sauber, gut dokumentiert und praxistauglich“*. Diese gute Dokumentation und Praxistauglichkeit werden vier Mal betont. Bei Jonas wird dieser Hinweis in keinem der zehn Schreiben verwendet. Hier wird wiederum das Phänomen verdeutlicht, dass Frauen in der IT häufig in sogenannten „Office Housework“-Tätigkeiten wie Dokumentieren und Organisieren tätig werden, die kaum karrierefördernd sind (Marsden et al. 2023).

Interessante Variationen ergeben sich auch in der Beschreibung besonders hervorzuhebender Eigenschaften und des Verhaltens gegenüber Kolleginnen und Kollegen. Bei Jonas wird vor allem auf die Beschreibung der Arbeitsweise Wert gelegt: Sein *„aktives“* Verhalten (4)¹⁰, seine *„Gewissenhaftigkeit“* (3) und *„Problemlösungsfähigkeit“* (2) und das *„Einbringen von Verbesserungspotenzialen“* (2). Bei Anna werden eher allgemeingültige Formulierungen gewählt, die sich auch bei Jonas finden. Darüber hinaus werden ihre sozialen Kompetenzen sprachlich deutlich ausgearbeitet. Sie *„überzeugt durch ein gutes Gespür für Zusammenarbeit in den Gruppen“*, hat *„eine hohe Sozialkompetenz“* und *„eine ausgeprägte Teamfähigkeit und Kommunikationsstärke“* und *„ist ein kollegiales Wesen“*. Doppelt so häufig wie Jonas war sie *„bei Kolleginnen und Kollegen sehr geschätzt“* (sechs zu drei Nennungen).

Ein letzter sehr deutlicher Unterschied zeigt sich am Ende der Schreiben, in dem formuliert wird, wofür Anna und Jonas empfohlen werden. Jonas wird an drei Stellen für eine Festanstellung empfohlen, an weiteren drei für *„eine verantwortungsvolle Position“*. Anna dagegen wird an fünf Stellen als *„wertvolle Mitarbeiterin“* beschrieben. Eine Festanstellung oder verantwortungsvolle Position sind in keinem der zehn Empfehlungsschreiben für sie vorgesehen.

3.3 Auswirkungen der Verzerrungen auf die Gleichstellung

Die beiden Forschungsexperimente zeigen die Subtilität der Ungleichbehandlungen durch ChatGPT: Die Erfolgsgeschichten wiesen ein ausgewogenes Geschlechterverhältnis auf, die Empfehlungsschreiben folgen dem gleichen Aufbau und sind standardisiert-wohlwollend formuliert. Dennoch zeigen die Details, dass ChatGPT zufolge Frauen seltener für Führungspositionen vorgesehen sind. Ihr Beitrag zum Erfolg wird lediglich in einer unterstützenden Mitarbeit und der Stärkung des sozialen Zusammenhalts gesehen.

Die Ergebnisse lassen vermuten, dass bei der Weiterentwicklung der Sprachmodelle bereits Anpassungen zum Ausgleich von Verzerrungen stattgefunden haben. Andere Beispiele aus der Literatur zu offensichtlicherer Geschlechterdiskriminierung von ChatGPT lassen sich nicht mehr reproduzieren. Der Versuch des zweiten Experiments, den neutralen, branchen- und positions-

10 Zahlen nach den Worten verweisen auf die Anzahl der Nennungen in den Empfehlungsschreiben.

unspezifischen Prompt für berufliche Empfehlungsschreiben von Wan et al. (2023) zu replizieren, scheiterte, da nahezu identische Schreiben erzeugt wurden. Erst die Anpassung auf die IT-Branche machte die beschriebenen Unterschiede zwischen den Schreiben sichtbar.

Auch wenn es bei der Modellentwicklung bereits Verbesserungen hinsichtlich stereotypischer Verzerrungen gab, sind latente Diskriminierungen weiterhin vorhanden und kritisch einzustufen. Dadurch, dass sie unterschwellig sind und nicht direkt auffallen, ist die Gefahr der Manipulierbarkeit umso größer. Folgende Beispiele verdeutlichen die Auswirkungen des Einsatzes von KI auf den Arbeitsmarkterfolg und die Gleichstellung von Frauen: Bei der Berufsberatung kann es durch den Einsatz von KI-Systemen zur Reproduzierung von Geschlechterstereotypen kommen, mit dem Ergebnis, dass Jungen eher Berufe im technischen Bereich und Mädchen Berufe im sozialen Bereich empfohlen werden (Raudonat et al. 2022). Auch bei der Stellensuche werden Frauen benachteiligt, wenn ihnen bei Jobbörsen schlechter bezahlte Jobs (Datta et al. 2015) oder bei Google seltener Werbeanzeigen für MINT-Stellen angezeigt werden (Lambrecht und Tucker 2018). Fortgeführt wird die Benachteiligung bei der Auswahl von Kandidatinnen und Kandidaten, wie weiter oben das Beispiel des namhaften Konzerns zeigte, ebenso wie eine Studie, die signifikant niedrigere Gehaltsvorschläge durch Sprachmodelle bei Frauen belegte (Sorokovikova et al. 2025).

4. Gendergerechte KI und Chancen für die Gleichstellung

4.1 Wie können KI-Systeme gendergerecht gestaltet werden?

Immer präsenter werden Begriffe wie „Faire KI“ oder „Feministische KI“. Bei diesen Konzepten steht das Ziel im Vordergrund, Vorurteile zu quantifizieren und die Diskriminierung von Untergruppen abzuschwächen (Feuerriegel et al. 2020). Dabei müssen verschiedene ethische Normen und Fairnesskonzepte in Erwägung gezogen werden, insbesondere wenn es um die Belange intersektionaler Gruppen geht (Himmelreich et al. 2025).

Um den Kreislauf der Diskriminierung zu unterbrechen, haben Marsden et al. (2023) ein Rahmenmodell zur Gestaltung sozialverantwortlicher Künstlicher Intelligenz entwickelt. Das Modell setzt an vier Interventionsmöglichkeiten an: Bei jeder Entwicklungsstufe einer KI-Anwendung sollte über die Sozialverträglichkeit reflektiert werden (**Prozess**). Dafür spielt Diversität in Entwicklungsteams eine wichtige Rolle, ebenso wie Schulungen zur Sensibilisierung im Bereich Gendergerechtigkeit (**Personen**). Vielfältige Trainingsdaten, inklusive Bedienoberfläche und Transparenz können helfen, Diskriminierungen zu vermeiden (**Produkt**). Grundlegende Anforderungen dazu legt der AI-Act der EU fest (siehe Infobox „Der rechtliche Rahmen der EU für Künstliche Intelligenz“). Relevant ist zudem der frühzeitige Einbezug der Anwendenden des Produkts, um Probleme aufzudecken (**Gesellschaft**). Durch das Modell sollen Strukturen und Vorgehensweisen so gestaltet werden, dass gleichberechtigte Teilhabe für alle gewährleistet wird (Marsden et al. 2023). Inwiefern die gesellschaftliche Mitgestaltung von KI möglich ist, zeigt das Praxisbeispiel AlgorithmWatch.¹¹

11 AlgorithmWatch (<https://algorithmwatch.org/de/>) ist eine Nichtregierungsorganisation aus Berlin und Zürich, deren Ziel es ist, gesellschaftliche und rechtliche Rahmenbedingungen so zu gestalten, dass KI und Algorithmen Gerechtigkeit, Demokratie, Menschenrechte und Nachhaltigkeit stärken.



Der rechtliche Rahmen der EU

2024 verabschiedete die EU mit dem „AI Act“ das weltweit erste Gesetz zur Regulation von KI. Es setzt grundlegende Standards für die Kontrolle über digitale Infrastrukturen und Daten und die Entwicklung menschenzentrierter und vertrauenswürdiger KI. Die Verordnung schafft einen einheitlichen, risikobasierten Rahmen für alle EU-Länder: Je höher das Risiko von KI-Anwendungen, desto strenger die Anforderungen und desto höher die Strafen:

- Minimales Risiko (zum Beispiel Spamfilter): keine besonderen Verpflichtungen.
- Begrenztes Risiko (zum Beispiel ChatBots, Bilderstellungstools): Transparenzverpflichtung gegenüber den Nutzenden.
- Hohes Risiko für die Gesundheit, Sicherheit oder Grundrechte (zum Beispiel Kreditwürdigkeitsbewertung): strenge Anforderungen an Datensätze, Risikobewertung, Dokumentation und Information.
- Verbotene Systeme: inakzeptables Risiko wie Manipulation oder unrechtmäßige Überwachung.

Die meisten Vorschriften gelten ab August 2026. Verstöße führen zu hohen Geldstrafen und Reputationsverlust.

Quelle: Verordnung (EU) 2024/1689 zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz (Verordnung über künstliche Intelligenz).

Praxisbeispiel: Interview mit Pia Sombetzki¹² von AlgorithmWatch

FaFo: Sie setzen sich dafür ein, dass KI und Algorithmen der ganzen Gesellschaft zugute kommen. Wie gehen Sie konkret vor?

Pia Sombetzki: Wir schauen beim Einsatz von KI ganz genau hin, welche Risiken algorithmische Systeme mit sich bringen, und wer die Verantwortung übernimmt. Wir wollen, dass möglichst viele Menschen verstehen, warum es sie ganz persönlich betrifft, wie KI eingesetzt wird. Deshalb informieren wir die Öffentlichkeit durch eigene Recherchen, wissenschaftliche Analysen, politische Einordnungen, Kampagnen und Informationsangebote und Medienarbeit. Außerdem entwickeln wir konkrete Lösungsempfehlungen für Politik, Entwicklerinnen und Entwickler und Betreiber von KI-Systemen.

Was konnten Sie durch Ihre Arbeit mit Blick auf die Diskriminierung der Geschlechter durch KI bereits erreichen?

Wir machen sichtbar, wie KI geschlechtsspezifische Diskriminierung reproduzieren kann und schaffen die Grundlage dafür, dass solche Systeme verändert werden. Dazu machen wir Fälle öffentlich und unterstützen Betroffene mit Informationen und Werkzeugen. Zum Beispiel:

12 Pia Sombetzki ist Policy Managerin für den Bereich algorithmische Diskriminierung.



Eine unserer Recherchen zeigt, dass Facebook bei Stellenausschreibungen Geschlechter-Stereotype verstärkt, wobei Jobs für LKW-Fahrerinnen und -Fahrern vor allem Männern und Anzeigen für Kinderbetreuung überwiegend Frauen angezeigt wurden. Solche Mechanismen können den Arbeitsmarkt massiv beeinflussen. Wir informieren zudem über externe Recherchen, etwa über eine Bloomberg-Analyse zum KI-Bildgenerator Stable Diffusion, der Frauen fast nur in schlecht bezahlten Jobs darstellte, als Kassiererin oder Haushälterin. So verbreiten sich Stereotype. Und wir haben ein Meldeformular¹³ entwickelt, über das sich Betroffene direkt an uns wenden können, wenn sie glauben, von Entscheidungen eines automatisierten Systems diskriminiert worden zu sein. Dadurch können sie sich gegen ungerechtfertigte Entscheidungen wehren und wir sammeln gleichzeitig Belege, mit denen wir überprüfen können, ob die Regulierung von KI-Anwendungen in der Praxis funktioniert.

Wie ist es möglich, Einfluss auf die faire Entwicklung von KI zu nehmen?

Wir haben durchaus Einfluss. Zum einen setzt die EU mit der KI-Verordnung verbindliche Standards, beispielsweise zur Transparenz und Risikobewertung. Diese gelten für alle Anbieter, die ihre Produkte hier in Europa auf den Markt bringen. Zum anderen spielt öffentlicher Druck eine wichtige Rolle. Zivilgesellschaft, Wissenschaft und Medien können Missstände aufzeigen und so Unternehmen und die Politik zu Maßnahmen bewegen. Wir selbst haben zum Beispiel untersucht, welche Verzerrungen Sprachmodelle enthalten, und aufgedeckt, dass sie falsche Wahlinformationen verbreiten. Microsoft hat daraufhin Schutzmaßnahmen verbessert.

Worauf können Unternehmen und Organisationen für eine faire Nutzung von KI achten?

Sie sollten sich vor dem Einsatz von KI immer fragen, welche Folgen auftreten können, diese dokumentieren und dabei Verantwortlichkeiten festlegen. Häufig werden bereits für Diskriminierung anfällige analoge Prozesse eins zu eins in automatisierte Prozesse überführt. Mittlerweile gibt es zahlreiche Tools, um Auswirkungen auf die Grundrechte von Menschen vorausezusehen. Auch wir haben so ein Tool entwickelt: Mit unserem kostenfreien Fragebogen können Unternehmen und Organisationen einen Transparenzbericht über den Einsatz ihrer KI-Systeme erstellen.¹⁴

Was können Nutzerinnen und Nutzer tun?

Sie sollten den Einsatz von KI kritisch hinterfragen und Diskussionen rund um den Einsatz von KI in der Gesellschaft verfolgen, um Risiken besser einschätzen zu können und neues Wissen zu teilen. Wer glaubt, durch einen Algorithmus diskriminiert worden zu sein, kann den Fall über unser Meldeformular melden. Generell ist es wichtig, Beweise zu sichern und Screenshots, E-Mails oder andere Unterlagen aufzubewahren. Wird beispielsweise ein Kredit überraschend schnell abgelehnt, kann die einzelne Person überlegen, welche Daten sie bei Antragsstellung angegeben hat, die vielleicht zu einer diskriminierenden Entscheidung geführt haben könnten.

13 <https://algorithmwatch.org/de/algorithmische-diskriminierung-melden/>, abgerufen am 08.09.2025.

14 <https://algorithmwatch.tech/impactassessment/>, abgerufen am 08.09.2025.

4.2 Chancen durch KI für die Gleichstellung

Bei allen negativen Konsequenzen von KI-Systemen haben diese durchaus auch Potenzial, Diskriminierungsdimensionen zu erkennen und positive Effekte auf die Gleichstellung von Frauen und Männern zu bewirken und somit zu einer gerechteren Gesellschaft beizutragen. Dies verdeutlichen die folgenden Beispiele:

Im Bereich der **Gesundheitsfürsorge** können KI-Systeme helfen, geschlechtsspezifische Unterschiede besser zu verstehen und zu behandeln. KI gestützte Programme tragen beispielsweise zur frühzeitigen Diagnose von Brustkrebs bei oder helfen, Risikoindikatoren für Herzinfarkte geschlechtsspezifisch zu identifizieren und so die Diagnostik zu verbessern (Kumar und Navysri 2023).

KI-Systeme können zudem zu mehr **Sicherheit und Schutz** von Frauen im Netz beitragen. Die KI-basierte App PENEMUE¹⁵ überprüft Kommentare und Nachrichten im Internet auf Anfeindungen. Die automatische Hatespeech-Erkennung ermöglicht es Anwenderinnen, unangemessene Inhalte schnell zu filtern und entsprechend darauf zu reagieren (zum Beispiel löschen). Gleichzeitig sammelt die App strafbare Inhalte und erleichtert so eine juristische Vorgehensweise. Insbesondere für Frauen, die in der Öffentlichkeit stehen, kann eine solche KI-Anwendung hilfreich sein¹⁶.

Eine App zur Unterstützung von Gründerinnen bietet KITE II¹⁷. Die KI-gestützte, gamifizierte Anwendung soll sie dabei unterstützen, resilienter im Umgang mit diskriminierenden Erfahrungen im Gründungsprozess zu werden. Dabei wurde ein innovatives Verfahren zur Trainingsdurchführung entwickelt, das sicherstellt, dass die zugrundeliegenden Algorithmen der Mustererkennung nicht nur diskriminierungsfrei sind, sondern auch Diskriminierung aufspüren und in Trainings integriert werden können. Eine Weiterentwicklung der App ist angedacht.

Auch in Unternehmensstrukturen können KI-Anwendungen Gleichstellung fördern. So entstehen durch deren Einführung **neue Möglichkeitsräume für Frauen**. Die Implementierung und Nutzung von KI-Systemen erfordert Kompetenzen aus den Bereichen Sprache, Ethik, Geistes-/Sozialwissenschaften und damit Studienfächern mit hohen Frauenquoten.¹⁸ So werden neue Rollen und Chancen für diverse und interdisziplinäre Teams geschaffen (Marsden 2024). Dies wiederum kann geschlechterbezogene Machtungleichheiten in betrieblichen Kontexten verändern und **Gleichstellungsbemühungen** fördern. KI-Anwendungen wie FAIR_solution¹⁹ unterstützen solche Prozesse. FAIR_solution erkennt vorhandene strukturelle Unterschiede zwischen den Geschlechtern (Gender Gaps) in Unternehmen und analysiert die Ungleichbehandlungen in den Bereichen Gehalt, Karriere, Arbeitszeit und Weiterbildung entlang der Diversitätsdimensionen Gender, Behinderung, Herkunft und Alter.

15 <https://www.penemue.ai/> (abgerufen am 08.09.2025), kladde, the creators GmbH, Freiburg.

16 Beispielsweise für Kommunalpolitikerinnen, die immer häufiger von digitaler Gewalt und Hatespeech betroffen sind und kaum Ressourcen zur Gegenwehr haben (FaFo 2023).

17 <https://www.kite-bga.de/> (abgerufen am 09.08.2025), Hochschule Heilbronn in Kooperation mit bundesweite gründerinnenagentur (bga).

18 <https://www.statistik-bw.de/BildungKultur/Hochschulen/>, abgerufen am 08.09.2025.

19 <https://www.ines-analytics.com/fair-solution> (abgerufen am 09.08.2025), INES Analytics.

5. Fazit

KI-Anwendungen erleichtern immer mehr Menschen den Alltag in verschiedensten Lebensbereichen. Der Report zeigte zudem, dass es bereits heute mit Unterstützung von KI gelingt, diskriminierende Strukturen zu erkennen und abzubauen. Ebenso wurde deutlich, dass jede und jeder Einzelne Einfluss auf die KI-Entwicklung nehmen kann. Dennoch reiht sich dieser Report in die Literatur nationaler und internationaler Studien ein, die zeigen, dass Künstliche Intelligenzen vorhandene soziale Ungleichheiten, Diskriminierungen und Stereotype reproduzieren, manifestieren und durch exponentielle Weiterverbreitung von Nutzenden skalieren. Bemerkenswert ist die beobachtete Tendenz, dass Verzerrungen subtiler werden und sich eher in den Details der Inhalte wiederfinden. Je subtiler die Ungleichbehandlung, desto schwieriger ist sie zu erkennen, zu benennen und auszuräumen. Umso wichtiger ist es, die Nutzenden generativer KI in ihrer KI-Kompetenz zu stärken.

Es gibt bereits einige Initiativen, die über KI aufklären und die Kompetenz von Nutzenden stärken. Frauennetzwerke wie The Female AI Club²⁰ können einen Safe Space für Frauen bieten, die KI verstehen und nachhaltig anwenden wollen. Interaktive Lernformate oder Buddy-Programme unterstützen die Nutzung und Mitentwicklung von KI-Anwendungen. Zur Stärkung der KI-Kompetenz von Frauen gehört auch die frühe Förderung von MINT-Interessen. Dies beginnt bei der Unterstützung von Mädchen und Frauen in ihrem Interesse an MINT-Fächern durch Eltern und Bildungspersonal (Martinot et al. 2025) und setzt sich fort in Programmen und Kursen für Mädchen, in denen sie Berührungspunkte mit IT und KI und deren Entwicklung erhalten.

Öffentliche Aufklärungskampagnen (wie von AlgorithmWatch) und Bildungsangebote können den kritischen und kompetenten Umgang mit KI-Technologien fördern. Entsprechende Schulungsangebote von verschiedenen Anbietern für unterschiedliche Zielgruppen finden sich beispielsweise auf der Lernplattform KI-Campus²¹.

Auch für Unternehmen und Institutionen wären Pflichtkurse im Bereich KI-Ethik vor der Einführung von KI-Produkten eine Möglichkeit, um Mitarbeitende für einen offenen aber zugleich kritischen Umgang mit KI zu schulen. Qualifizierungsangebote bietet beispielsweise das Verbundprojekt RIZ AI TRAQC (Artificial Intelligence Training & Qualification Campus)²².

Um die Potenziale von KI-Anwendungen positiv für die Gesellschaft nutzen zu können, ist es wichtig, soziale Innovationen und gemeinwohlorientierte KI-Projekte zu unterstützen (zum Beispiel Civic Coding – Innovationsnetzwerk KI für das Gemeinwohl)²³. In Hinblick auf Gleichstellungsfragen ist die Förderung von KI-Anwendungen zu betonen, die diskriminierende Strukturen aufdecken und gleichstellungspolitische Ziele unterstützen können.

20 <https://www.thefemaleaiclub.com/>, abgerufen am 08.09.2025.

21 <https://ki-campus.org/> (abgerufen am 09.08.2025), Stifterverband für die Deutsche Wissenschaft e.V.

22 <https://www.hs-heilbronn.de/de/artificial-intelligence-training-and-qualification-campus-3b52bc6f3e823a64>, abgerufen am 08.09.2025.

23 <https://www.civic-coding.de/>, abgerufen am 08.09.2025.



Eine zukunftsfähige und verantwortungsvolle KI-Entwicklung erfordert divers besetzte Teams, kritische digitale Kompetenz in allen Bevölkerungsgruppen und den klaren gesellschaftlichen und politischen Willen, Technologie als Werkzeug für die Teilhabe aller Menschen zu nutzen. Der Schutz der Grundrechte aller Bürgerinnen und Bürger, auch im digitalen Bereich, ist ein zentraler Grundsatz der europäischen KI-Verordnung vom 13. Juni 2024, in der die EU weltweit erstmalig verbindliche Regelungen für den Einsatz von KI festlegt. Das bedeutet, dass unter anderem das Grundrecht der Gleichstellung der Geschlechter die Basis für weitere KI-Entwicklungen und die digitale Zukunft in Baden-Württemberg bildet. Es besteht Handlungsbedarf, sicherzustellen, dass KI – auch in ihren lernenden und oft subtilen Ausprägungen – keine Benachteiligung von Frauen zur Folge hat. Die Landesregierung hat in ihrer Gleichstellungsstrategie konkrete Maßnahmen zu KI im Bereich Diskriminierung und Gewaltschutz entwickelt. Dazu zählt die Förderung eines Projekts, das zeitnah beginnen wird. Innerhalb des Projekts werden die Diskriminierung von Frauen durch KI beim Personalrecruiting untersucht und konkrete Schutzstrategien gegen digitale Gewalt und Diskriminierung am Arbeitsplatz erarbeitet.

6. Literatur

Anwar, Usman; Saparov, Abulhair; Rando Ramirez, Javier; Paleka, Daniel; Turpin, Miles; Hase, Peter; Lubana, Ekdeep Singh; Jenner, Erik; Casper, Stephen; Sourbut, Oliver; Edelman, Benjamin L.; Zhang, Zhaowei; Günther, Mario; Korinek, Anton; Hernandez-Orallo, Jose; Hammond, Lewis et al. (2024): Foundational Challenges in Assuring Alignment and Safety of Large Language Models. ETH Zürich.

Arntz, Melanie; Baum, Myriam; Brüll, Eduard; Dorau, Ralf; Hartwig, Matthias; Lehmer, Florian; Matthes, Britta; Meyer, Sophie-Charlotte; Schlenker, Oliver; Tisch, Anita; Wischniewski, Sascha (2025a): Digitalisierung und Wandel der Beschäftigung (DiWaBe 2.0): Eine Datengrundlage für die Erforschung von Künstlicher Intelligenz und anderer Technologien in der Arbeitswelt. Baua: Bericht.

Arntz, Melanie; Brüll, Eduard; Burkert, Carola; Matthes, Britta (2025b): Abgehängt? Frauen nutzen KI beruflich viel seltener als Männer. In: IAB-Forum. Online unter: <https://iab-forum.de/graphs/abgehaengt-frauen-nutzen-ki-beruflich-viel-seltener-als-maenner/>, abgerufen am 15.08.2025.

Barera, Michael (2020): „Mind the Gap: Addressing Structural Equity and Inclusion on Wikipedia“. Librarian & UTA Libraries Staff Publications, 112.

Barthel, Michael; Stocking, Galen; Holcomb, Jesse; Mitchel, Amy (2016): Nearly Eight-in-Ten Reddit Users Get News on the Site. Pew Research Center.

Chatterji, Aaron; Cunningham, Thomas; Deming, David J.; Hitzig, Zoe; Ong, Christopher; Shan, Carl; Wadmann, Kevin (2025): How People Use ChatGPT. National Bureau of Economic Research (NBER). Working Paper 34255.

Datta, Amit; Datta, Anupam; Tschantz, Michael Carl (2015): Automated Experiments on Ad Privacy Settings: A Tale of Opacity, Choice, and Discrimination. In: Proceedings on Privacy Enhancing Technologies, 1, S. 92–112.

Eloundou, Tyna; Manning, Sam; Mishkin, Pamela; Rock, Daniel (2023): Gpts are gpts: An early look at the labor market impact potential of large language models.

FaFo (2023): Hatespeech gegen Kommunalpolitikerinnen in Baden-Württemberg. GesellschaftsReport BW, Ausgabe 2 – 2023.

Ferran-Ferrer, Núria; Boté-Vericad, Juan-José; Minguillón, Julià (2023): “Wikipedia gender gap: a scoping review”. In: Profesional de la información, 32, 6, e320617.

Feuerriegel, Stefan; Dolata, Mateusz; Schwabe, Gerhard (2020): Fair AI. In: Business & Information Systems Engineering, 62, S. 379–384.

Frau und Beruf Baden-Württemberg (2022): Daten & Fakten zu Frau und Beruf. Online unter: <https://www.frauundberuf-bw.de/zahlen-daten-fakten-frau-und-beruf>, abgerufen am 15.08.2025.

Gengler, Eva; Kraus, Andreas, Bodrožić-Brnić, Kristina (2024): Faires KI-Prompting – Ein Leitfaden für Unternehmen. In: BSP Business and Law School – Hochschule für Management und Recht, 1–28.

Haensch, Anna-Carolina; Ball, Sarah; Herklotz, Markus; Kreuter, Frauke (2023): Seeing ChatGPT through students' eyes: An Analysis of TikTok Data.

Hargittai, Eszter (2020): Potential biases in big data: Omitted voices on social media. In: Social Science Computer Review, 38, 1, S. 10–24.

Himmelreich, Johannes; Hsu, Arbie, Veomett, Ellen; Lum, Kristian (2025): The Intersectionality Problem for Algorithmic Fairness. In: Proceedings of Machine Learning Research, 279, S. 68–95.

Hunkenschroer, Anna Lena; Luetge, Christoph (2022): Ethics of AI-Enabled Recruiting and Selection: A Review and Research Agenda. In: Journal of Business Ethics, 178, S. 977–1007.

Initiative D21 e. V. (2025): D21-Digital-Index 2024/25. Jährliches Lagebild zur Digitalen Gesellschaft. Online unter: <https://initiated21.de/publikationen/d21-digital-index/2024-25>, abgerufen am 15.08.2025.

JIM-Studie (2024): Jugend, Information, Medien. Basisuntersuchung zum Medienumgang 12- bis 19-Jähriger. Medienpädagogischer Forschungsverbund Südwest c/o Landesanstalt für Kommunikation (LFK).

Koenig, Anne. M.; Eagly, Alice H.; Mitchell, Abigail. A.; Ristikari, Tiina (2011): Are leader stereotypes masculine? A meta-analysis of three research paradigms. In: Psychological Bulletin, 137, 4, S. 616–642.

Kumar, Rayi Naveen; Navyasri, Mullapudi (2023): Predicting the Heart Attacks Risk Using Artificial Neural Networks. Paper presented at the International Conference on Science, Technology and Engineering.

Lambrecht, Anja; Tucker, Catherine E. (2018): Algorithmic Bias? An Empirical Study into Apparent Gender-Based Discrimination in the Display of STEM Career Ads. In: Management Science, 65, 7.

Marsden, Nicola; Raudonat, Kerstin; Pröbster, Monika (2023): Kreislauf der Diskriminierung. In: Frauen machen Informatik. Magazin der GI-Fachgruppe Frauen in der Informatik, 47, 2023.

Marsden, Nicola (2024): Sozialverantwortliche KI-Gestaltung. Online-Kurs. KI-Campus.

Martinot, P.; Colnet, B.; Breda, T.; Sultan, J.; Touitou, L.; Huguet, P.; Spelke, E.; Dehanae-Lambertz, G.; Bressoux, P.; Dehaene, S. (2025): Rapid emergence of a maths gender gap in first grade. In: *Nature*, 643,8073, S. 1020–1029.

Mosene, Katharina; Rachinger, Felicitas (2025): KI zwischen technologischem Fortschritt und gesellschaftlicher Verantwortung. In: *blog interdisziplinäre geschlechterforschung*, 22.04.2025. Online unter: <http://www.gender-blog.de/beitrag/ki-fortschritt-und-verantwortung/> abgerufen am 12.09.2025.

Nieken, Petra (2025): Nicht im Datensatz, nicht in der Welt. Karlsruhe: KIT. Online unter: <https://www.peba.kit.edu/6797.php>, abgerufen am 01.09.2025.

Orwat, Carsten (2019): Diskriminierungsrisiken durch Verwendung von Algorithmen. Antidiskriminierungsstelle des Bundes.

Ouyang, Long; Wu, Jeff; Jiang, Xu; Almeida, Diogo; Wainwright, Carroll L.; Mishkin, Pamela; Zhang, Chong; Agarwal, Sandhini; Slama, Katarina; Ray, Alex; Schulman, John; Hilton, Jacob; Kelton, Fraser; Miller, Luke; Simens, Maddie; Askell, Amanda; Welinder, Peter; Christiano, Paul; Leike, Jan; Lowe, Ryan (2022): Training language models to follow instructions with human feedback. OpenAI. Online unter: https://cdn.openai.com/papers/Training_language_models_to_follow_instructions_with_human_feedback.pdf, abgerufen am 15.09.2025.

Raudonat, Kerstin; Pröbster, Monika; Schmieder, Kai; Martinetz, Simone; Marsden, Nicola (2022): Where bias can creep in – Gendersensibilität beim Einsatz von KI-Technologien im Kontext beruflicher Weiterbildung im Forschungsprojekt KIRA. In: Demmler, Daniel, Daniel Krupka und Hannes Federrath (Hrsg.): *Informatik 2022, Lecture Notes in Informatics (LNI)*, Gesellschaft für Informatik, Bonn. S. 683–692.

Sadok, Hicham; Sakka, Fadi; El Hadi El Maknouz, Mohammed (2022): Artificial intelligence and bank credit analysis: A review. In: *Cogent Economics & Finance*, 10, 22, 1.

Schwaber, Ken; Sutherland, Jeff (2020): *The Scrum Guide. The Definitive Guide to Scrum: The Rules of the Game*. Online unter: <https://scrumguides.org/index.html>, abgerufen am 15.09.2025.

Sorokovikova, Aleksandra; Pavel Chizhov, Iuliia Eremenko und Ivan P. Yamshchikov (2025): Surface Fairness, Deep Bias: A Comparative Study of Bias in Language Models. In: *Proceedings of the 6th Workshop on Gender Bias in Natural Language Processing (GeBNLP)*. S. 206–227.

Statistisches Landesamt Baden-Württemberg (2024): 2023: 31 % mehr weibliche MINT-Beschäftigte in Baden-Württemberg als 2013. Frauen in MINT-Berufen jedoch weiter unterrepräsentiert. Pressemitteilung 51/2024.



Statistisches Landesamt Baden-Württemberg (2025): Immer mehr Frauen entscheiden sich für ein MINT-Studium. Zum Internationalen Frauentag am 8. März 2025. Pressemitteilung 56/2025.

Tanlamai, Jitsama; Khern-am-nuai, Warut; Cohen, Maxime C. (2025): Generative AI and Price Discrimination in the Housing Market. Online unter: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4764418, abgerufen am 15.09.2025.

Tiggemann, Marika; Anderberg, Isabella (2020): Social media is not real: ‚Instagram vs reality‘ images on women’s social comparison and body image. In: New media & society 22, 12, S. 2183–2199.

Wan, Yixin; Pu, George; Sun, Jiao; Garimella, Aparna; Chang, Kai-Wei; Peng Nanyun (2023): “Kelly is a Warm Person, Joseph is a Role Model”: Gender Biases in LLM-Generated Reference Letters. In: Findings of the Association for Computational Linguistics: EMNLP 2023, S. 3730–3748.

World Economic Forum (2019): Global Gender Gap Report 2020. Online unter: https://www3.weforum.org/docs/WEF_GGGR_2020.pdf, abgerufen am 15.08.2025.



Impressum

Der GesellschaftsReport BW wird herausgegeben vom
Ministerium für Soziales, Gesundheit und Integration
Baden-Württemberg
Else-Josenhans-Straße 6
70173 Stuttgart
Tel.: 0711 123-0
Internet: <https://sozialministerium.baden-wuerttemberg.de/>

Autorinnen

Stephanie Bundel, Dr. Felizia Hanemann

Redaktion

Helin Hoffmann

Layout

Andrea Mohr

Copyright-Hinweise

© Statistisches Landesamt Baden-Württemberg, Fellbach, 2025

Fotonachweis Titelbild

Scusi / Fotolia

Verteilerhinweis

Diese Informationsschrift wird von der Landesregierung in Baden-Württemberg im Rahmen ihrer verfassungsmäßigen Verpflichtung zur Unterrichtung der Öffentlichkeit herausgegeben. Sie darf weder von Parteien noch von deren Kandidatinnen und Kandidaten oder Helferinnen und Helfern während eines Wahlkampfes zum Zwecke der Wahlwerbung verwendet werden. Dies gilt für alle Wahlen.

Missbräuchlich ist insbesondere die Verteilung auf Wahlveranstaltungen, an Informationsständen der Parteien sowie das Einlegen, Aufdrucken und Aufkleben parteipolitischer Informationen oder Werbemittel.

Untersagt ist auch die Weitergabe an Dritte zum Zwecke der Wahlwerbung. Auch ohne zeitlichen Bezug zu einer bevorstehenden Wahl darf die vorliegende Druckschrift nicht so verwendet werden, dass dies als Parteinahme des Herausgebers zugunsten einzelner politischer Gruppen verstanden werden könnte. Diese Beschränkungen gelten unabhängig vom Vertriebsweg, also unabhängig davon, auf welchem Wege und in welcher Anzahl diese Informationsschrift dem Empfänger zugegangen ist.

Erlaubt ist es jedoch den Parteien, diese Informationsschrift zur Unterrichtung ihrer Mitglieder zu verwenden.