

# KI VERSTEHEN?

## IN KÜRZE

- Künstliche Intelligenz (KI) ist bereits fixer Bestandteil vieler Alltagsanwendungen, von der Suchmaschine bis hin zur Kreditvergabe.
- Wie Entscheidungen von KI getroffen werden, ist häufig unklar.
- Transparenz ist wichtig, um die Nachvollziehbarkeit von KI-Entscheidungen zu erhöhen.
- Um Transparenz zu gewährleisten, sind technische Möglichkeiten allein nicht ausreichend.
- KI muss im sozialen Kontext gesehen werden und Regulierung soziale Auswirkungen von KI-Anwendungen stärker in den Blick nehmen.

## WORUM GEHT ES?

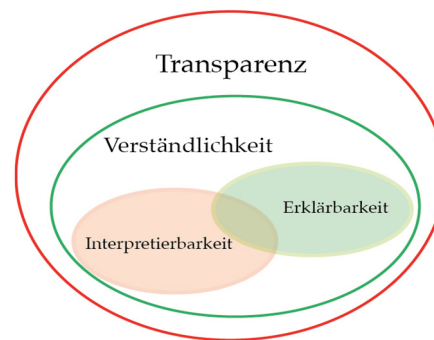
„Künstliche Intelligenz“ (KI) ist allgegenwärtig. Großflächige Anwendungen von KI basieren auf umfassenderen und genaueren Datenanalysen. Das weckt Hoffnungen auf eine effizientere, objektivere Gestaltung von wirtschaftlichen oder sonstigen gesellschaftlichen Prozessen. Aber was sind die Konsequenzen und Nebeneffekte eines breiten Einsatzes? Entscheidungen, die durch KI gefällt werden, bleiben häufig im Dunkeln und sind teilweise sogar für Entwickler\*innen unerklärlich („black box“).

Um die Anwendung von KI sicherer zu gestalten und Auswirkungen auf den Menschen abzuschätzen, wird höhere Transparenz bzw. Nachvollziehbarkeit gefordert. Diese Forderung bezieht sich aber meist auf Entwickler\*innen, Produzent\*innen und den kommerziellen Vertrieb, während potenziell (Mit-)Betroffene wie Konsument\*innen oder andere Nutzer\*innen oft unbeachtet bleiben. Des Weiteren gibt es wesentliche Hindernisse: Es ist weder eindeutig definiert, was KI alles miteinschließt, noch, was mit Transparenz genau gemeint ist.

KI vereint verschiedene gesellschaftsrelevante Ansprüche, dient aber auch als Sammelbecken für unterschiedliche technische Ansätze wie Algorithmen, maschinelles Lernen oder algorithmische Entscheidungssysteme. Transparenz wiederum kann auf den Algorithmus, den Prozess oder den Kontext bezogen sein. Sie kann technisch, z.B. als die detaillierte Offenlegung von Codes,

oder prozedural, z.B. als zielgerichtete Kommunikation zu bestimmten Zielgruppen, verstanden werden. Daher reicht eine Forderung nach Transparenz allein nicht aus, um KI verantwortungsvoll zu entwickeln, in Umlauf zu bringen und zu regulieren.

Das Forschungsfeld der erklärbaren KI (Explainable AI, XAI) setzt es sich zum Ziel, KI-Entscheidungen nachvollziehbarer zu machen. Es versucht, Transparenz durch technische externe Verständnishilfen bzw. Schnittstellen herzustellen und außerhalb des KI-Systems zu kommunizieren sowie Erklärungen über Entscheidungen in die Entwicklung von Algorithmen einzubeziehen. Zu den Merkmalen einer solchen Schnittstelle gehören etwa Datenvisualisierung oder Szenarioanalysen.



Dimensionen von Transparenz

Quelle: nach Clinciu/Hastie (2019)

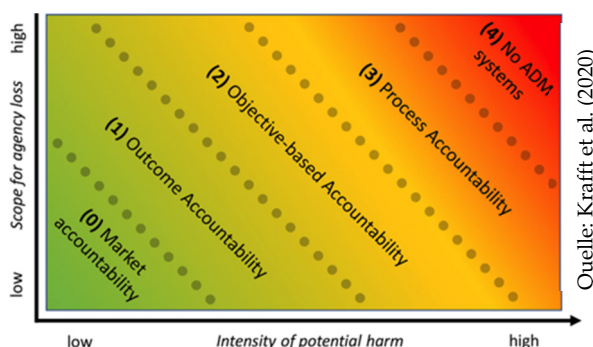
Für eine verantwortungsvolle KI-Entwicklung und wirksame Regulierung bleibt die Frage, ob solche Ansätze ausreichen, um soziale Nachvollziehbarkeit und Verantwortung zu ermöglichen und damit soziale und ethische Probleme zu entschärfen. Rein technische Umsetzungen von Transparenz erlauben keine Aussagen über soziale Auswirkungen von KI-Systemen. Daher kann sich Regulierung von KI nicht ausschließlich über technische Aspekte definieren, sondern muss Auswirkungen von KI auf den Menschen in den Mittelpunkt des Interesses rücken.

## ECKDATEN

Projekttitel:	Künstliche Intelligenz – Verstehbarkeit und Transparenz
Projektteam:	Udrea, T., Fuchs, D., Peissl, W.
Laufzeit:	10/2021 – 01/2022
Auftraggeber:	Bundesarbeitskammer

## WESENTLICHE ERGEBNISSE

Schon jetzt verändern sich Sozialsysteme durch IT- und KI-Anwendungen, die Entscheidungen unterstützen (z.B. bei der Jobsuche, der Zuteilung von Sozialleistungen oder wenn es um Kreditwürdigkeit geht). Ein breiter Einsatz von KI-Systemen zieht weitere Veränderungen nach sich. Bereits mit herkömmlicher IT können gesellschaftliche Probleme entstehen, wie z.B. Diskriminierung durch statistische Verfahren. Der Einsatz von KI könnte solche Probleme noch verschärfen. Wie Transparenz konkret gestaltet werden soll, diskutiert man in der Forschung und in der Politik intensiv. Fest steht: Transparenz oder Erklärbarkeit dürfen nicht nur auf technischer Ebene umgesetzt werden, sondern müssen auch für von KI Betroffene bzw. Interessensvertretungen nachvollziehbar und überprüfbar gestaltet werden.



### Eine mögliche Einteilung von KI-Systemen nach Risikostufen

Daher sollten regulatorische Ansätze neben technisch-basierten Zugängen auch und vor allem gesellschaftliche Anwendungskontexte berücksichtigen. Unterschiedlich risikoreiche KI-Anwendungen könnten in einem System abgestufter Transparenzregeln reflektiert werden. Anwendung können so von unabhängigen Institutionen kontrolliert werden, die auch Betroffene bei Transparenzforderungen, Beschwerden oder Klagen aktiv unterstützen können. Da die Funktionsweisen von KI-Anwendungen nicht immer vollständig abgeschätzt werden können, wäre es sinnvoll, nur solche Systeme einzusetzen, die theoretisch von Menschen kontrolliert werden können. Im Vordergrund sollte in jedem Fall die Verantwortung gegenüber den von KI-Anwendungen Betroffenen stehen, sei es durch regulierende Stellen, durch Entwickler\*innen oder durch Produzent\*innen. Betroffene und Interessensgruppen müssen aktiv in Gestaltungsprozesse von KI eingebunden werden, um deren Sozialverträglichkeit zu erhöhen.

## WAS TUN?

Um das Potenzial von KI auszuschöpfen und eine verantwortungsvolle und sozialverträgliche Gestaltung voranzubringen, erscheint es notwendig,

- Interessen von Konsument\*innen und Betroffenen im KI-Diskurs und in der Entwicklung zu stärken,
- einen ausdifferenzierten Kriterienkatalog zur Klassifizierung von KI und ihren Risiken zu erarbeiten,
- eine KI-Definition für Regulierungen zu verwenden, die auch etablierte IT-Systeme mit entsprechenden Auswirkungen auf den Menschen umfasst,
- Transparenz in ihren verschiedenen Dimensionen – wie etwa dem Recht auf Information, als Nachvollziehbarkeit für Menschen und als institutionell verankerte Transparenz inklusive der notwendigen Verfahren (Rechtsbehelfe) – umzusetzen,
- KI-Systeme, deren Entscheidungen Menschen betreffen, generell zu registrieren und nach erwartetem Risiko zu zertifizieren,
- einen Regulierungsansatz zu verfolgen, der periodisch evaluiert wird,
- Forschung zu erklärbarer KI (XAI), Fairness, Gerechtigkeit, Rechenschaftspflicht, Verantwortlichkeit und gesellschaftlichen Auswirkungen zu fördern,
- ethisch bedenkliche KI-Systeme oder solche, die Grundrechte, Freiheiten oder die Demokratie beschädigen, zu verbieten und
- menschliche Kontrolle über KI-Systeme zu gewährleisten und als Grundanforderung zu stärken. Solange dies nicht möglich ist, sollten Moratorien für derartige KI-Systeme überlegt werden.

## ZUM WEITERLESEN

Udrea, T., Fuchs, D., Peissl, W. (2022) Künstliche Intelligenz – Verstehbarkeit und Transparenz. ITA-Projektbericht 2022-01. in Kooperation mit der Bundesarbeitskammer.  
[pub.oeaw.ac.at/ita/ita-projektberichte/ITA-2022-01.pdf](http://pub.oeaw.ac.at/ita/ita-projektberichte/ITA-2022-01.pdf)

## KONTAKT

Walter Peissl  
E-Mail: [tamail@oeaw.ac.at](mailto:tamail@oeaw.ac.at)  
Telefon: +43 1 51581-6582

