Vertrauenswürdige KI – Ethik-Richtlinien der EU-Kommission

Ethik und Künstliche Intelligenz? Geht das zusammen? Ist nicht schon längst entschieden, dass allein der Markt bestimmt, was machbar und erlaubt ist? Nein, denn mit dem KI-Einsatz in der Industrie ist die Frage der Ethik wieder in den Vordergrund getreten.

In den letzten zwei bis drei Jahrzehnten haben wir uns daran gewöhnt, dass sich Digitalisierung und mit ihr Künstliche Intelligenz in immer mehr Bereiche der Konsumwelt ausgedehnt haben, ohne dass sich irgendein Staat um die Frage gekümmert hat, ob das mit rechten Dingen zugeht, ob Ethik und Moral unserer Gesellschaft, die Freiheit des Individuums und der demokratische Rechtsstaat dabei auch nur im Blick sind. Oft genug war dies nicht der Fall.

Erst das Internet der Dinge und der beginnende KI-Einsatz in der Industrie und in anderen Bereichen der Wirtschaft und Gesellschaft hat das Thema Ethik nun auf die Tagesordnung gebracht, und zwar zuallererst in Europa, wo bereits mit der Datenschutzgrundverordnung (DSGVO) ein Pfahl für einen verantwortungsvolleren Umgang mit personenbezogenen Daten in den Boden gerammt wurde.

Die erste Ethik-Kommission für Künstliche Intelligenz wurde vom Bundesministerium für Verkehr und digitale Infrastruktur (BMVI) zum Thema „Automatisiertes und Vernetztes Fahren“ eingerichtet und legte im Juni 2017 ihren Bericht vor. Im April 2019 wurden dann von einer High-Level Expert Group zur Künstlichen Intelligenz (AI HLEG), die die EU-Kommission im Sommer 2018 zusammengerufen hatte, einerseits eine Definition vertrauenswürdiger KI veröffentlicht, andererseits Ethik-Richtlinien für die Entwicklung und Nutzung vertrauenswürdiger KI-Systeme. (Das im August 2020 bei Hanser erscheinende Buch „KI-Kompass für Entscheider“ behandelt diese Richtlinien in Kapitel 5.4.)

Die rund 40 Seiten füllenden Richtlinien erheben den Anspruch, Entwickler und Nutzer von KI-Systemen dabei zu unterstützen, ethischen Grundprinzipien gerecht zu werden. Die Richtlinien stellen drei Grundanforderungen an vertrauenswürdige KI:

- Sie sollte gesetzeskonform sein und allen jeweils zutreffenden Gesetzen und Regulierungen folgen.
- Sie sollte sich an ethischen Prinzipien und Werten ausrichten.
- Sie sollte zuverlässig sein, und zwar sowohl in technischer als auch in sozialer Hinsicht, denn selbst mit bester Absicht entwickelte KI kann unbeabsichtigten Schaden verursachen.

Auch wenn die Richtlinien lediglich die Vorstellungen einer Expertengruppe sind und nicht etwa den Rang von Regulierungen oder EU-Gesetzen haben, sind sie als Richtschnur zu begrüßen. Es war höchste Zeit, dass ethische Rahmen definiert werden, in denen die weitere Entwicklung der KI an eine gesellschaftlich wünschenswerte Leine gelegt wird.

Die vier Grundprinzipien der europäischen Union werden zitiert und zur generellen Grundlage der KI-Ethik erklärt: Respekt der menschlichen Selbstbestimmung, Schadensvorbeugung, Fairness und Erklärbarkeit. Damit stellen die Richtlinien die menschliche Autonomie absolut in den Vordergrund. Sich selbst steuernde Systeme, die die Entscheidungsgewalt und -verantwortung des Menschen ersetzen, sind folglich nicht unter vertrauenswürdiger KI einzuordnen.

Systeme der Künstlichen Intelligenz sollen nach diesen Richtlinien die besondere Verwundbarkeit bestimmter Gruppen – beispielsweise Kinder, Behinderte und andere immer wieder ausgegrenzte

Menschen – berücksichtigen. Der Asymmetrie der Macht zwischen Arm und Reich, Unternehmern und abhängig Beschäftigten, Wirtschaft und Verbrauchern sollen KI-Systeme entgegenwirken, statt sie zu verstärken. Das beruht auf der Erfahrung, dass etliche derzeit eingesetzte KI-Systeme das genau entgegengesetzte Ziel zu verfolgen scheinen. So können Stellenbewerbungen oder Wohnungsgesuche regelrecht rassistisch oder frauenfeindlich vorsortiert und gefiltert sein, je nachdem, wie die verwendete KI trainiert wurde. Systeme der Künstlichen Intelligenz sollten deshalb, so die Richtlinien aus Brüssel, stets darauf überprüfbar sein, ob sie diesen ethischen Prinzipien gerecht werden.

Der zweite Teil der Richtlinien erläutert sieben konkrete Anforderungen, die vertrauenswürdige KI erfüllen soll:

1. Menschliches Handeln, die Grundrechte und die letztlich menschliche Beaufsichtigung müssen gewährleistet sein.
2. KI muss technische Robustheit und Sicherheit bieten.
3. Datenschutz und Datenverwaltung gelten auch und erst recht für KI.
4. Die Funktion von KI-Systemen und insbesondere ihre jeweilige Entscheidungsfindung sollen transparent und rückverfolgbar sein.
5. KI-Systeme sollen Vielfalt, Nichtdiskriminierung und Fairness dienen.
6. Das ökologische und gesellschaftliche Wohlergehen muss bei KI-Systemen höchste Priorität haben.
7. Ergebnisse des KI-Einsatzes müssen überprüfbar sein, negative Auswirkungen sollen minimiert, Rechtsmittel und Verantwortbarkeit sichergestellt sein.

Für die Einhaltung dieser Prinzipien sollen mit der Entwicklung und dem Angebot von KI-Systemen auch technische und nichttechnische Methoden bereitgestellt werden.

Im dritten Teil der Richtlinien zeigen praktische Vorschläge, wie eine Bewertungsliste bei der Entwicklung und Nutzung vertrauenswürdiger KI-Systeme helfen kann. Und schließlich sind einige sehr konkrete Beispiele aufgeführt, die entweder den besonderen Nutzen der KI-Anwendung demonstrieren, etwa im Gesundheitswesen, oder den dramatischen Schaden, wenn es um autonome, tödliche Waffensysteme geht.

Die Richtlinien sind kein europäisches Gesetz und keine Grundordnung wie die DSGVO. Aber sie sollten einer breiteren Öffentlichkeit bekannt gemacht werden, denn sie können dabei helfen, aus europäischen KI-Systemen Vorbilder wachsen zu lassen. „Die Einhaltung der Richtlinien soll, so der von den Experten formulierte Wunsch in der Einleitung, nicht die Einführung von KI behindern, sondern sie soll selbst zu einem Markenzeichen werden, das weltweit den wirtschaftlichen Erfolg vertrauenswürdiger Systeme deutlich steigert.“ (Ulrich Sendler, KI-Kompass für Entscheider, Hanser Verlag, München, ISBN: 978-3-446-46295-3, erscheint im August 2020, Kapitel 5.4)