

Vertrauenswürdigkeit von Anwendungen der Künstlichen Intelligenz – Konsequenzen für den Agrarsektor

Reiner Brunsch¹

Abstract: Im Beitrag werden, ausgehend von der (in Europa genutzten) Definition Künstlicher Intelligenz, die Komponenten und Anforderungen von vertrauenswürdiger Künstlicher Intelligenz (vKI) dargestellt, wie sie von der Europäischen Kommission empfohlen werden. Da KI nicht Selbstzweck ist, sondern einen Beitrag zur Mehrung des Gemeinwohls leisten soll, gilt es diese Effekte künftig zu bewerten, neben der Erfüllung der Anforderungen über den gesamten Lebenszyklus. Die Frage der Vertrauenswürdigkeit bezieht sich nach Auffassung von Experten und Europäischer Kommission nicht nur auf die KI an sich, sondern auch auf die zugrundeliegenden Daten und die verfolgten Geschäftsmodelle. Es wird aufgezeigt, wie in anderen Teilen der Welt mit ethischen Grundsätzen der KI-Anwendung umgegangen wird. Neben den Möglichkeiten zur Bewertung der Vertrauenswürdigkeit werden Spezifika der Agrarproduktion vorgestellt, die eine sektorspezifische Bewertung begründen.

Keywords: Künstliche Intelligenz, Vertrauenswürdigkeit, Bewertungsmethoden, Agrarproduktion

1 Einleitung

Über die Funktionen von Künstlicher Intelligenz (KI) ist in jüngster Zeit international teils sehr kontrovers diskutiert und entschieden worden. Dabei hat das Thema Vertrauenswürdigkeit zunehmend an Bedeutung gewonnen. Der Begriff „Künstliche Intelligenz“ wird hier im Sinne der Definition der Hochrangigen Expertengruppe für Künstliche Intelligenz (HEG-KI) verwendet [Et19]. Die folgenden Aussagen zu KI-Systemen beziehen sich prinzipiell auf deren gesamten Lebenszyklus. Die hohe Komplexität der Agrarproduktion macht diese besonders attraktiv für KI-Anwendungen.

2 Vertrauenswürdigkeit von KI-Anwendungen „Made in Europe“

Die Europäische Kommission setzt sich dafür ein, den Aspekt der Vertrauenswürdigkeit als wesentliches Charakteristikum europäischer Entwicklungen zu unterstützen [Wh20]. In den EU-Ethik-Richtlinien für eine vertrauenswürdige KI wird darauf verwiesen, dass die KI kein Selbstzweck ist, sondern ein vielversprechendes Mittel, um das menschliche Gedeihen und somit das Wohlbefinden von Individuum und Gesellschaft und das Gemeinwohl zu steigern sowie zur Förderung von Fortschritt und Innovation beizutragen

¹ Leibniz-Institut für Agrartechnik und Bioökonomie, Max-Eyth-Allee 100, 14469 Potsdam, rbrunsch@atb-potsdam.de,

[Et19]. Allgemeiner auf den gesamten Prozess der Digitalisierung bezogen, wird vom Wissenschaftlichen Beirat der deutschen Bundesregierung für globale Umweltveränderungen (WBGU) in der Zusammenfassung seines Hauptgutachtens 2019 die Erwartung formuliert, dass „die Digitalisierung so gestaltet werden muss, dass sie als Hebel und Unterstützung für die Große Transformation zur Nachhaltigkeit dienen und mit ihr synchronisiert werden kann“ [Un19].

Eine vertrauenswürdige KI zeichnet sich laut HEG-KI [Et19] durch *drei Komponenten* aus, die während des gesamten Lebenszyklus des Systems erfüllt sein sollten: a) Sie sollte *rechtmäßig* sein und somit alle anwendbaren Gesetze und Bestimmungen einhalten, b) sie sollte *ethisch* sein und somit die Einhaltung ethischer Grundsätze und Werte garantieren und c) sie sollte *robust* sein, und zwar sowohl in technischer als auch sozialer Hinsicht, da KI-Systeme selbst bei guten Absichten unbeabsichtigten Schaden anrichten können. Jede Komponente an sich ist notwendig, jedoch nicht ausreichend, um das Ziel einer vertrauenswürdigen KI zu erreichen. Idealerweise wirken alle drei Komponenten harmonisch zusammen und überlappen sich in ihrer Funktionsweise.

Es muss gewährleistet sein, dass die Entwicklung, Einführung und Nutzung von KI-Systemen die *Anforderungen* an vertrauenswürdige KI erfüllen: 1) Vorrang menschlichen Handelns und menschliche Aufsicht, 2) technische Robustheit und Sicherheit, 3) Schutz der Privatsphäre und Datenqualitätsmanagement, 4) Transparenz, 5) Vielfalt, Nichtdiskriminierung und Fairness, 6) gesellschaftliches und ökologisches Wohlergehen sowie 7) Rechenschaftspflicht. Alle Anforderungen sind in Bezug auf ihre Bedeutung gleichrangig. Die Umsetzung dieser Anforderungen sollte für den gesamten Lebenszyklus eines KI-Systems erfolgen und sich nach der spezifischen Anwendung richten. Zwar treffen die meisten Anforderungen auf sämtliche KI-Systeme zu, ein besonderes Augenmerk muss jedoch auf diejenigen gelegt werden, welche direkt oder indirekt Menschen betreffen.

Da sich KI-Systeme in einem dynamischen Umfeld ständig weiterentwickeln, ist die Schaffung einer vertrauenswürdigen KI ein kontinuierlicher Prozess, in dessen Zentrum die Bewertung steht (siehe Abschnitt 4).

3 Internationale Positionen zur Vertrauenswürdigkeit von KI

Die **OECD**-Grundsätze [OE19] verlangen eine künstliche Intelligenz (KI), die innovativ und vertrauenswürdig ist und die Menschenrechte und demokratische Werte achtet. Sie wurden am 22. Mai 2019 von den OECD-Mitgliedsländern angenommen. Die KI-Grundsätze der OECD sind die ersten von Regierungen unterzeichneten Grundsätze dieser Art.

Basierend auf den KI-Grundsätzen der OECD ist das Thema auch ein Handlungsfeld der Vereinigung der **G20-Staaten**. Auf einer Fachministerkonferenz im Juni 2019 kam es zu

einer Übereinkunft [G219]. Die Zusammenfassung eines Multi-Stakeholder-Dialogs zur Rolle der Vertrauenswürdigkeit im Gesundheits- und Bildungskontext enthält u.a. Empfehlungen an die Regierungen, nicht nur das Vertrauen in KI-Anwendungen durch geeignete Maßnahmen zu stärken, sondern auch das Monitoring über die KI-Anwendungen und deren Wirkungsanalyse zu beschleunigen [Su20].

Die OECD listet für die **USA** derzeit 40 nationale Initiativen für KI auf und wird als globale Nummer 1 in der Forschung zu KI gesehen [AI20]. Es fällt in den nationalen Veröffentlichungen auf, dass eher von Verantwortung (responsibility) und ethischer KI (ethic AI) geschrieben wird. Es scheint so, als würden Gesellschaft und IT-Wirtschaft zunehmend erkennen, dass KI nur eine Chance hat, wenn sie auch die Menschen mitnimmt. Einen Überblick über die aktuellen Debatten und Wertungen geben PANG [Pa20] und KIND [Ki20]. Zum Jahresende 2019 forderte HAO dazu auf, das „AI ethics-washing“ zu stoppen und wirklich etwas zu tun [Ha19]. Dies charakterisiert den Beginn der dritten Welle (KIND), in der KI als sozio-technisches System verstanden wird und auch entsprechende Instrumente zum Audit und der Bewertung entwickelt werden.

China hat als G20-Mitglied die KI-relevanten Dokumente mitgezeichnet. Die Regierung hat in Juni 2019 die Grundsätze für die neue Phase des Umgangs mit KI veröffentlicht. Diese heben das Thema der Entwicklung verantwortungsvoller künstlicher Intelligenz hervor und betonen acht Prinzipien: Harmonie und Freundschaft, Fairness und Gerechtigkeit, inklusives Teilen, Achtung der Privatsphäre, Sicherheit und Kontrollierbarkeit, gemeinsame Verantwortung, offene Zusammenarbeit und agile Governance [Pr19].

4 Bewertungsmethoden für vertrauenswürdige KI

Die Europäische Kommission hat im Juli 2020 die Bewertungsliste für vertrauenswürdige KI veröffentlicht [Th20]. Sie ist Ergebnis der Arbeit der HEG-KI, für die Eigenbewertung vorgesehen und orientiert sich an den sieben Anforderungen an eine vKI. Somit besteht für die EU-Staaten ein einheitliches Bewertungsinstrumentarium.

Prinzipiell sind Standards und Normen nützliche Instrumente für die Überwachung und Zertifizierung technischer Prozesse, die in der Regel Ergebnis einer demokratischen Aushandlung sind [Be20a]. Unter der Annahme, dass es sich bei vKI um ein sozio-technisches System handelt, würde der Prozess der Standardentwicklung sich nicht nur auf die technischen Aspekte beziehen können [Be20a]. Unter Berücksichtigung der besonderen Herausforderung schlagen TIEDRICH und QADIR [TQ20] für die Operationalisierung von vKI den Weg über „best practices“ vor.

Die Kompetenzplattform KI.NRW fördert ein Projekt zum Aufbau eines Zertifizierungsrahmens für KI-Systeme [Wh19].

Eine weitere Frage der Bewertung betrifft den Nutzen der KI-Anwendung für die Menschheit. Dafür steht bisher kein allgemein anerkanntes Verfahren zur Verfügung. Einen Ansatz können die z.B. zivilgesellschaftlichen Diskurse liefern, bei denen es beispielsweise um die Frage geht, ob Algorithmen das Gemeinwohl fördern können [Be20b]. Andere Ansätze sind im April 2020 veröffentlicht worden [Ex20]. In diesem Report wird die Bewertung unterteilt in das Audit der Algorithmen und die Wirkungsanalyse der Algorithmen, wobei bei der Wirkungsanalyse zwischen Risikoabschätzung (vor der Anwendung) und Bewertung der Auswirkungen (während der Anwendung) unterschieden wird.

5 Gründe für eine sektorspezifische Anpassung der Bewertungsmethoden in der Agrarwirtschaft

Bereits in der Ethik-Richtlinie der HEG-KI [Et19] wird auf die eventuelle Notwendigkeit einer sektorspezifischen Anpassung der Richtlinie hingewiesen.

Die Besonderheit der landwirtschaftlichen Primärproduktion im Vergleich zu anderen Wirtschaftsbereichen wird seit längerem aus verschiedenen Sichtweisen hervorgehoben. So wird die Agrarproduktion beispielsweise anhand folgender Merkmale charakterisiert [Pe05]: Es handelt sich um agrarisch geprägte Ökosysteme, die mit sozio-technischen und sozio-ökonomischen Um- und Supersystemen in Beziehungen stehen. Das Agrarökosystem kann noch in seine Subsysteme Boden, Pflanzen, Tiere unterteilt werden, die miteinander interagieren und von Menschen und Maschinen beeinflusst werden. Diese zentralen Subsysteme der Agrarproduktion sind biologische Systeme, die einerseits über die Fähigkeit der Selbstregulation verfügen, aber gleichzeitig gegenüber ihrer Umwelt als weitgehend offene Systeme funktionieren. Deshalb werden Agrarproduktionssysteme im Kontext der Digitalisierung als cyber-physical-biological-social-ecological systems (cpbses) gesehen, in denen 5 Welten aufeinander treffen: die physische, biologische, humane/soziale, die Informationswelt und die Umwelt [Br19]. Im Unterschied zu Industrie 4.0, wo von cyber-physical systems (cps) gesprochen wird, sind bei KI-Anwendungen im Agrarbereich einige Dimensionen mehr zu berücksichtigen.

Eine weitere Besonderheit ist die moralische und rechtliche Rolle der in der Nutztierhaltung Beschäftigten. Tierschutz ist seit 2002 in Deutschland Staatsziel. Niemand darf einem Tier ohne vernünftigen Grund Schmerzen, Leiden oder Schäden zufügen. Die Rechtsposition der Tiere ist detailliert und aktuell beschrieben, u.a. in [DT19]. Bezüglich der Auswirkungen für KI-Anwendungen wurden bisher keine wissenschaftlichen Veröffentlichungen gefunden.

Mit der Digitalisierung wird zwischen den Menschen und den anderen Elementen eines Agrarökosystems eine neue Ebene eingezogen, der „digitale Vorhang“. Insbesondere der

komplexer Informationsaustausch zwischen Menschen und anderen biologischen Systemen wird verändert, im ungünstigen Fall behindert. vKI muss dies berücksichtigen.

6 Ausblick

Damit Vertrauenswürdigkeit das Kennzeichen von (europäischen) KI-Anwendungen in der Landwirtschaft werden kann, braucht es eine Berücksichtigung der Besonderheiten der Agrarproduktion. Ziel einer vom Autor initiierten (europäischen) Arbeitsgruppe ist es, diese Besonderheiten für den gesamten Lebensprozess von KI-Systemen zu berücksichtigen und daraus einen Bewertungsrahmen zu entwickeln, der wissenschaftlich begründet ist und in einem breiten Stakeholder-Diskurs zur Akzeptanz gebracht werden soll. Auf diese Weise soll die technologische Innovation KI zu einem akzeptierten, gemeinwohlorientierten sozio-technischen System im Agrarbereich werden. Als übergeordnete Ziele werden die Gewährleistung von Ernährungssicherheit und Ernährungssouveränität im Digitalzeitalter angesehen.

Literaturverzeichnis

- [AI20] OECD (2020b) AI in United States <https://www.oecd.ai/dashboards/countries/UnitedStates>. Abruf: 15.11.2020.
- [Be20a] BEINING, L. (2020a) Vertrauenswürdige KI durch Standards? <https://www.stiftung-nv.de/sites/default/files/herausforderungen-standardisierung-ki.pdf>. Abruf: 15.11.2020.
- [Be20b] BEINING, L. (2020b) Können Algorithmen das Gemeinwohl stärken? <https://www.stiftung-nv.de/de/publikation/koennen-algorithmen-das-gemeinwohl-staerken>. Abruf: 15.11.2020.
- [Br19] BRUNSCH, R. (2019) Stellungnahme zu „Chancen und Risiken der Digitalisierung in der Landwirtschaft“, Deutscher Bundestag, https://www.bundestag.de/resource/blob/592038/f9a405974cd67de52d7dc25ed64011a/f/Stn_von_Prof_Brunsch-data.pdf. Abruf: 15.11.2020.
- [DT19] DIEHL, E. und TUIDER, J. (2019) Haben Tiere Rechte? <https://www.bpb.de/shop/buecher/schriftenreihe/297138/haben-tiere-rechte>. Abruf: 15.11.2020.
- [Et19] HEG-KI (2019): Ethik-Leitlinie für eine vertrauenswürdige KI, April 2019, <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai> Abruf: 15.11.2020.
- [Ex20] ALI (2020) Examining the Black Box: Tools for Assessing Algorithmic Systems, Ada Lovelace Institute, <https://www.adalovelaceinstitute.org/examining-the-black-box-tools-for-assessing-algorithmic-systems/> Abruf:15.11.2020.
- [G219] G20 (2019) G20 Ministerial Statement on Trade and Digital Economy, Annex: G20 AI Principles, June 2019 <https://www.mofa.go.jp/files/000486596.pdf>. Abruf: 15.11.2020.

- [Ha19] HAO, K. (2019) In 2020, let's stop AI ethics-washing and actually do something <https://www.technologyreview.com/2019/12/27/57/ai-ethics-washing-time-to-act/> Abruf: 15.11.2020.
- [Ki20] KIND, C. (2020) The term 'ethical AI' is finally starting to mean something, Carly Kind, August 23, 2020 <https://venturebeat.com/2020/08/23/the-term-ethical-ai-is-finally-starting-to-mean-something/> Abruf: 15.11.2020.
- [OE19] OECD (2019a) OECD-Grundsätze für künstliche Intelligenz https://www.oecd.org/berlin/presse/Flyer_AIPrinciples_FINAL_GER.pdf Abruf: 14.11.2020.
- [Pa20] PANG, W. (2020) Responsible AI becomes critical in 2021 (Nov 2020) <https://venturebeat.com/2020/11/11/responsible-ai-becomes-critical-in-2021/> Abruf: 15.11.2020.
- [Pe05] DFG (2005) Perspektiven der agrarwissenschaftlichen Forschung, Denkschrift, Deutsche Forschungsgemeinschaft https://www.dfg.de/download/pdf/dfg_im_profil/geschaeftsstelle/publikationen/perspektiven_agrarwiss_forschung_de.pdf. Abruf: 15.11.2020.
- [Pr19] MOST (2019) Principles of next-generation AI governance - responsible AI http://most.gov.cn/kjbgz/201906/t20190617_147107.htm. Abruf: 14.11.2020.
- [Su20] G20 (2020): Summary of discussions from the G20 AI dialogue in 2020, <https://g20.org/en/media/Documents/Summary%20of%20Discussions%20from%20the%20G20%20AI%20Dialogue%20in%202020.pdf>. Abruf: 15.11.2020.
- [Th20] EC (2020b) The Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment <https://op.europa.eu/de/publication-detail/-/publication/73552fcd-f7c2-11ea-991b-01aa75ed71a1> Abruf: 15.11.2020.
- [TQ20] TIEDRICH, L. und QADIR, L. (2020) 10 Steps To Creating Trustworthy AI Applications <https://www.cov.com/-/media/files/corporate/publications/2020/04/10-steps-to-creating-trustworthy-ai-applications.pdf> Abruf: 15.11.2020.
- [Un19] WBGU (2019) Unsere gemeinsame digitale Zukunft, Hauptgutachten, Zusammenfassung https://issuu.com/wbgu/docs/wbgu_hgd2019_z. Abruf: 15.11.2020.
- [Wh19] IAIS (2019) Whitepaper-KI-Zertifizierung, Fraunhofer-IAIS, https://www.iais.fraunhofer.de/content/dam/iais/KINRW/Whitepaper_KI-Zertifizierung.pdf. Abruf: 15.11.2020.
- [Wh20] EC (2020a) White paper: On Artificial Intelligence - A European approach to excellence and trust https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf. Abruf: 15.11.2020.