

Aline Franzke und Sebastian Gießler

Vom Labor in die Gesellschaft: Model Cards als Werkzeug für verantwortungsvolle Robotik

Ein Leitfaden zur ethischen Gestaltung autonomer Systeme im Rahmen des Projekts EGAR: Ethische und gesellschaftliche Aspekte
Autonomer Robotik

20. Januar 2026

Dieser Leitfaden entstand im Rahmen des Forschungsprojekts:
“Ethische und gesellschaftliche Aspekte autonomer Robotik (EGAR)”

Unter der Leitung von Dr. Mone Spindler und Dr. Wulf Loh

Finanziert von der Baden-Württemberg Stiftung



Beim zunehmenden Einsatz von autonomen Robotern entstehen nicht nur für Anwender*innen, sondern auch für die Entwickler*innen ethische und gesellschaftliche Herausforderungen. Dieser Leitfaden soll Wissenschaftler*innen und Entwickler*innen in die Lage versetzen, ethische und soziale Aspekte autonomer Robotik bereits während des Projektverlaufs zu reflektieren und diesen Prozess kompetent zu gestalten.

Aufbau des Leitfadens

Nach einer kurzen Einführung in das Konzept von Model Cards und wie sich diese zu Ethikrichtlinien und Gesetzen verhalten, wird eine Reihe von Fragen zu ethischen und gesellschaftlichen Aspekten in der Robotikentwicklung vorgestellt. Diese Fragen werden idealerweise vor und während eines laufenden Forschungsprojekts bearbeitet. Sie dienen der systematischen Dokumentation, dem Wissensmanagement sowie der Vorbereitung von Publikationen, etwa zur Darstellung ethischer Aspekte des Forschungsprojekts in Konferenzbeiträgen oder Fachzeitschriften

Einführung

Model Cards wurden 2019 in dem Paper *Model Cards for Model Reporting* von Mitchell et al. entwickelt, und sollen die Eigenschaften, Kapazitäten und intendierte Nutzung von Machine-Learning-Systemen dokumentieren und eine Transparenz über die beabsichtigte Verwendung und technische Grenzen bieten. Ziel dieses Dokuments ist es, zum einen Model Cards als strukturierte Dokumentation vorzustellen, und mit einem Fragenkatalog um ethische und robotikspezifische Aspekte zu erweitern. Wir haben uns für Model Cards entschieden, da dieser Ansatz aus einer technischen Perspektive entwickelt wurde und wir uns von daher eine höhere Akzeptanz und eine bessere Integration in vorhandene Arbeitsprozesse und wissenschaftliche Praktiken versprechen. Dieser Ansatz ist explizit auch für Projekte anwendbar, die keine Machine-Learning-Ansätze verwenden.

Wir erweitern mit diesem Fragenkatalog Model Cards um Aspekte der Robotikentwicklung, die auf der Softwareebene nicht repräsentiert sind, etwa Interaktion mit Umwelt und Personen, und Roboterdesign und Körperlichkeit (*embodiment*). Wir verwenden diesen Ansatz, als ein Vorbild (Role Model) für andere computergestützte Technikwissenschaften, da Model Cards wesentliche Faktoren der

Softwareentwicklung im Machine-Learning Bereich berücksichtigen.

Was sind Model Cards? Der Model Card Ansatz kommt ursprünglich aus der Softwareentwicklung (*software engineering/SWE*) und wird vor allem in diesem Kontext eingesetzt. Sie sind als standardisierte begleitende Dokumentation konzipiert, die einen Bewertungsmaßstab für den geplanten Einsatzzweck verwendet werden kann.

Der Fokus liegt dabei auf klassischen Problemfeldern im Machine-Learning, etwa Details über das Modell, Datenherkunft, Verwendung von Trainingsdaten, intendierter Einsatzzweck eines Systems, vorgenommene Gewichtungen, und Limitierungen des Modells (Mitchell et al. 2019).

Model Cards und dieser Fragenkatalog sind freiwillige Dokumentationen und keine rechtlich bindenden Nachweispflichten. Sie sind ein sogenanntes *Soft Law*, also eine freiwillige Übereinkunft oder Absprache, im Gegensatz zu *Hard Law*, einer zwingenden Rechtsnorm.

Soft Law und *Hard Law* sind Begriffe der Corporate Governance. Beispiele für *Hard Law* sind etwa die EU-Maschinenverordnung (2023/1230/EU), das Medizinproduktegesetz (MPDG-2020) oder der EU AI Act (2024/1689/EU). Weitere Beispiele für *Soft Law* sind Ethikrichtlinien und freiwillige Selbstverpflichtungen von Institutionen oder Unternehmen.

Ein Nachteil des Model Card Ansatzes ist, dass die technische Ebene zwar strukturiert bearbeitet, die ethische Ebene jedoch unterentwickelt bleibt. „Ethical Considerations“ werden zwar explizit mit Beispielen vorgeschrieben, es wird aber keine strukturierte Anleitung, wie genau ethische Fragen zu bearbeiten sind, angeboten. (Mitchell et al. 2019: 6). In Anbetracht der globalen Fragmentierung von Ethics Governance und lokaler Eigenheiten von Forschungsinstitutionen ist es nachvollziehbar, sich nicht auf eine spezifische ethische Theorie oder ethische Prinzipien festzulegen.

Fragen zu ethischen Implikationen werden häufig an Ethikkommissionen, Institutional Review Boards (IRB) oder Research Ethics Boards (REB) verwiesen. Ein Problem dabei ist, dass diese Institutionen oftmals nicht für ethische Aspekte der Informatik oder der Ingenieurwissenschaften zuständig sind. Insbesondere in kleinen Organisationen, wie Unternehmen oder Hochschulen, fehlt es oft gänzlich an

einer entsprechenden Anlaufstelle. Dadurch sind Entwickler*innen häufig auf ihre individuelle Selbsteinschätzung angewiesen, was zu inkonsistenten oder unzureichenden ethischen Bewertungen führen kann.

Dieser Leitfaden zielt darauf ab, zusätzlich zur technischen Ebene die Reflexion von ethischen und gesellschaftlichen Aspekten zu strukturieren, um Entwickler*innen zu befähigen, sich selbstständig systematisch durch diese Aspekte zu arbeiten. Dieser Ansatz wurde von anderen Begleitprojekten bereits erfolgreich eingesetzt (McLennan et al. 2022; Willem et al. 2024)

Diese Systematik hat für die Projekte mehrere Vorteile:

*Robotikentwickler*innen.* Andere Entwickler*innen können anhand dieser Dokumentation nachvollziehen, wie ein technisches, soziales oder ethisches Problem adressiert wurde, und für die eigenen Problemlösungen als Referenz verwenden.

Policymaker. Politische Entscheidungsträger*innen können anhand der Dokumentation ethische und soziale Risiken abschätzen, und auf deren Grundlage entscheiden, ob Projekte gefördert werden sollten, Regulation nötig ist und ob soziotechnische Lösungsversprechen realistisch sind.

Organisationen. Mit einer Dokumentation können Organisationen, z.B. Unternehmen, entscheiden, ob sie ein bestimmtes System anschaffen wollen, wie sie es rechtssicher einsetzen können, und wie dieses System optimal in die unternehmenseigenen Strukturen eingebunden werden kann.

Betroffene Personen. Betroffene Personen können anhand der Dokumentation leichter nachvollziehen, welche Auswirkungen das Autonome System auf sie und ihre Lebenswelt haben könnte, und können fahrlässigen oder missbräuchlichen Einsatz leichter erkennen und kritisieren.

Ethische angepasste Model Cards für autonome Systeme

Die Dokumentation ist idealerweise ein einseitiger „Beipackzettel“, welcher für die relevanten Stakeholder, Förderorganisationen und andere Forscher*innen vorgehalten wird. Die Verschränkung von Model Cards mit ethischer Reflexion kann als ein Werkzeug für ethische Mitigation verstanden werden. Diese bietet für Roboterentwicklungsprojekte einen bedeutenden Mehrwert während der Design- und Entwicklungsphase. Internationale Zeitschriften und

Konferenzen setzen ebenso zunehmend Ethikevaluationen voraus.

Durch die Integration der ethischen Reflexion und der Dokumentation durch Model Cards adressieren wir ein weiteres, strukturelles, Problem: Im Vergleich zu anderen Ansätzen im Bereich der Technikfolgenabschätzung und ELSA (Ethical, legal, and societal aspects), wird der ethische Impuls durch die Entwickler*innen selbst durchgeführt und in die Projektpraktiken verankert: ethische und soziale Aspekte werden so in den Forschungsprozess integriert, und ELSA-Forscher*innen werden nicht erst am Ende des Projekts hinzugezogen (Post-hoc) und bewerten einen fertig entwickelten Demonstrator.

Model Cards und EGAR-Fragenkatalog

Dieser Fragenkatalog soll diese Lücke schließen und damit beide Ebenen verbinden – eine strukturierte technische Dokumentation und darauf aufbauend, eine strukturierte ethische Analyse und Risikobeurteilung. Um die Zugangshürde zu senken, kann dieser Fragenkatalog ohne vorheriges Erstellen einer Model Card bearbeitet werden. Einige Themen und Fragen kommen daher sowohl beim Model Card Ansatz als auch in diesem Fragenkatalog vor.

Der optimale Zeitpunkt für das Bearbeiten des EGAR-Fragenkatalogs ist, wenn bereits ein Grundkonzept erarbeitet und die Anforderungen vorliegen. So kann geklärt werden, *welche* Probleme *wie* mit einem Roboter gelöst werden sollen, bevor die verwendeten Technologien, die Datensätze und weitere Designentscheidungen eines Projekts final feststehen. Dies entspricht in der Regel dem ersten Drittel bis zur Hälfte des Entwicklungsprozesses.

Im Idealfall gibt es eine definierte Person oder Rolle im Entwicklungsteam, die für die ethische Bewertung zuständig ist. Dadurch wird sichergestellt, dass eine Person diese Themen bearbeitet und darauf achtet, dass Richtlinien eingehalten und umgesetzt werden. Dennoch bleiben ethische und soziale Fragestellungen eine Querschnittsaufgabe und sollten daher im gesamten Entwicklungsteam integriert werden.

Der Fragenkatalog hat zwei Funktionen: Fragen nach technischen Eigenschaften, Leistungsfähigkeit und intendiertem Einsatz autonomer Systeme werden überprüft und in ethisch-normative Argumente eingebettet. Beispielsweise führt die Frage, welche Datensätze verwendet wurden, zu Folgefragen, ob bei eigenen und fremden Datensätzen ethische und soziale Aspekte berücksichtigt wurden. Für die technische Ebene sind dabei die Model Cards nach



Mitchell et al. 2019 ein Vorbild¹, die für die domänenspezifischen Bedürfnisse der Robotikentwicklung erweitert werden. Diese Überlegungen richten sich an Entwickler*innen, Förderinstitutionen, und Forschungsgruppen, denen ethische und soziale Aspekte in der Forschungspraxis begegnen, und basieren auf einer Kombination von reflexiven und formalen Ansätzen der Technikfolgenabschätzung.

Reflexive Ansätze

Reflexive Ansätze konzentrieren sich auf die kritische Untersuchung der Überzeugungen, Annahmen und Praktiken von Forscher*innen und darauf, wie diese den Forschungsprozess beeinflussen. Reflexivität erfordert das Hinterfragen der eigenen Position und bewusster und unbewusster Überzeugungen des Forschers und die Anerkennung ihrer Rolle bei der Gestaltung des Forschungsprozesses. Reflexive Ansätze können Forscher*innen herausfordern, sowohl ihre persönlichen Perspektiven als auch die Annahmen innerhalb ihrer Disziplin zu untersuchen.

Diese Methoden basieren auf der Kommunikation und Praxis reflexiver Argumente in strukturierten Diskussionen. Diskussionen beruhen in der Regel auf einem dialogischen Konzept, das den Austausch zwischen ethischen Expert*innen (oder Werkzeugen) und Entwickler*innen fördert. Sie sind oft niedrigschwellig angelegt, um ethische Probleme sichtbar und intersubjektiv zugänglich zu machen. Gewählte Methoden sind in der Regel strukturierte oder teilstrukturierte Fragebögen, Workshops oder dialogische Zielgruppenanalysen und stützen sich auf interdisziplinäre Perspektiven, um ethische Aspekte zu bearbeiten.

Reflexive Ansätze sind oft mit Herausforderungen konfrontiert, da sie nur begrenzt operationalisierbar sind und von der Bereitschaft der Teilnehmenden zum Dialog abhängen. Sie sind zeitaufwändig und erfordern interdisziplinäre Arbeit, die eine gemeinsame Zielvorstellung und Fachsprache erfordert, um

komplexe technische und ethische Aspekte über verschiedene disziplinäre Hintergründe hinweg zu verstehen. Zudem ist fraglich, ob ethische Interventionen zu dauerhaften Veränderungen in den Prozessen führen können.

*Beispiel: Unser Prototyp funktioniert perfekt, warum weigern sich unsere Nutzer*innen, ihn zu benutzen?*

Formale Ansätze

Formale Ansätze bieten ein standardisiertes Verfahren, das oft aus den technischen Wissenschaften selbst stammt oder in Zusammenarbeit mit ihnen entwickelt wurde. Beispiele hierfür sind Ansätze wie Model Cards und Data Sheets. Diese Ansätze werden häufig aus der Informatik oder Ingenieurwissenschaft heraus entwickelt und basieren auf den üblichen Dokumentationsanforderungen technischer Systeme, d.h., sie erfordern wenig zusätzlichen Aufwand, da sie sich auf etablierte Praktiken in der Informatik und in der Ingenieurwissenschaft beziehen.

Dadurch sind sie in der Regel gut an die fachspezifischen Methoden, Prozesse und Vorgehensweisen dieser Disziplinen angepasst. Dadurch lassen sich diese formalen Ansätze reibungslos in die Arbeitsabläufe in Forschung und Entwicklung integrieren, weil sie von vornherein so konzipiert wurden. Im Vordergrund steht dabei die sachgerechte Dokumentation, d.h. Designentscheidungen, Entscheidungen über die verwendeten Datensätze, Datenherkunft, für welchen Zweck das System entwickelt wurde, und Details über die verwendeten statistischen und informatischen Methoden – ein „Beipackzettel“ für Technik.

Beispiel: Mein Prototyp funktioniert, wie dokumentiere ich Leistungsfähigkeit und Grenzen für die technische Fachcommunity?

¹ Vorbild im Sinne eines Role Models oder established protocol Ansatzes.



Literaturverzeichnis

McLennan, S., Fiske, A., Tigard, D., Müller, R., Haddadin, S., & Buyx, A. (2022). Embedded ethics: a proposal for integrating ethics into the development of medical AI. *BMC Medical Ethics*, 23(1), 6.

<https://doi.org/10.1186/s12910-022-00746-3>

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., et al. (2019). Model Cards for Model Reporting. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (pp. 220–229). <https://doi.org/10.1145/3287560.3287596>

Willem, T., Fritzsche, M.-C., Zimmermann, B. M., Sierawska, A., Breuer, S., Braun, M., et al. (2024). Embedded Ethics in Practice: A Toolbox for Integrating the Analysis of Ethical and Social Issues into Healthcare AI Research. *Science and Engineering Ethics*, 31(1), 3. <https://doi.org/10.1007/s11948-024-00523-y>