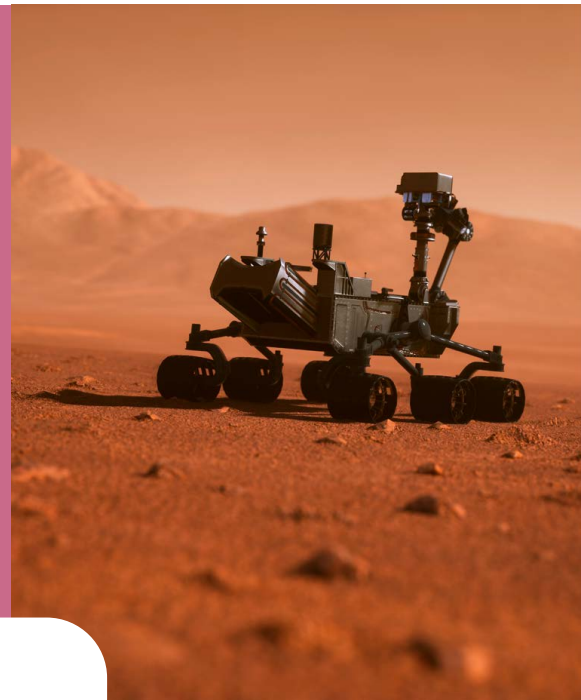


Kompetent im Einsatz

Variable Autonomie Lernender Systeme in lebensfeindlichen Umgebungen

Whitepaper von Jürgen Beyerer et al.
Arbeitsgruppe Lebensfeindliche Umgebungen



Kurzfassung

Egal ob im Weltraum, in der Tiefsee oder in Katastrophengebieten – Einsätze in solchen lebensfeindlichen Umgebungen stellen für den Menschen eine große Herausforderung mit erheblichen Risiken dar. Lernende Systeme können hier helfen, Gefahren und Risiken für Menschen zu verringern oder solche Umgebungen überhaupt erst zu erschließen. Die Missionskonfigurationen, der Grad der Autonomie eines selbstlernenden Systems sowie die Intensität der Interaktion mit Menschen können dabei enorm variieren. Entscheidend für das Gelingen der Zusammenarbeit zwischen Menschen und Lernenden Systemen ist eine gute Arbeitsteilung. Die Arbeitsgruppe Lebensfeindliche Umgebungen der Plattform Lernende Systeme hat Schlüsselvoraussetzungen für diese Arbeitsteilung zwischen Mensch und Lernendem System sowie für die Kompetenz des Lernenden Systems im jeweiligen situativen Anwendungskontext identifiziert und untersucht.

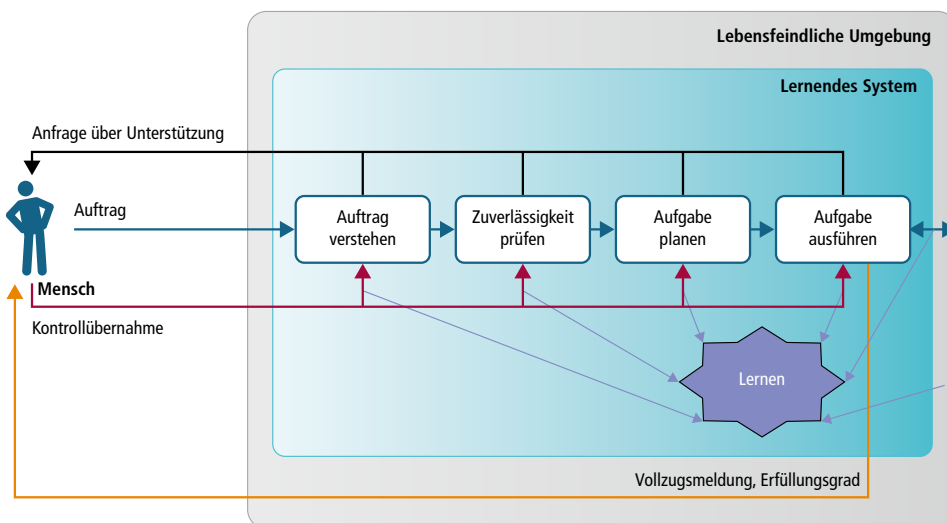
Lernende Systeme in lebensfeindlichen Umgebungen sind im Vergleich zu Anwendungsdomänen wie Industrie und Verkehr sehr individuell ausgestaltet. Zum gegenwärtigen Zeitpunkt können derartige Einsätze daher nicht ohne den Menschen als Überwacher des Geschehens konzipiert werden. Stattdessen geht es darum, den Menschen zu unterstützen und sein Gefahrenrisiko zu minimieren. Der Kerngedanke lautet: So viel Autonomie wie möglich – nur so viel menschlicher Eingriff wie nötig.

Das Whitepaper adressiert mit Bezug zu praxisorientierten Anwendungsfällen die Fragen, warum für Lernende Systeme in lebensfeindlichen Umgebungen eine variable Autonomie angestrebt werden sollte, welche Architekturkomponenten solche Systeme benötigen, welche Forschungsbedarfe existieren und welche Fragestellungen sich daraus für Anwendungen ergeben.

Grundlagen für die Zusammenarbeit: Mensch und Lernende Systeme

Für den Einsatz Lernender Systeme in lebensfeindlichen Umgebungen werden zunächst die spezifischen Anforderungen für den Grad der Autonomie solcher Systeme definiert. Dieser Autonomiegrad ist unter anderem abhängig von der Umgebung und der Art der Aufgabe, der Fähigkeit des Lernenden Systems sowie rechtlicher und ethischer Richtlinien für dessen Einsatz. Dieser muss darüber hinaus so angepasst werden, dass Menschen nur dort eingreifen müssen, wo es nötig und sinnvoll ist. Zudem müssen die Systeme aufgrund der Dynamik von Einsätzen in lebensfeindlichen Umgebungen einerseits in der Lage sein, den Grad ihrer Autonomie an die jeweilige Situation selbst anzupassen, oder andererseits ggf. durch den Menschen angepasst werden; dies bedarf einer Erweiterung bisheriger Konzeptionen von Autonomiegraden um die Perspektive variierbarer Autonomie während der Einsatzzeit eines Lernenden Systems. Das Lernende System ist immer als Mensch-Maschine-System zu verstehen, das heißt, es gibt eine Interaktion mit dem Menschen, die sehr unmittelbar sein kann (z. B. durch Fernsteuerung) oder nur punktuell erfolgt (z. B. durch Anpassung von Missionen).

Vereinfachter Workflow eines autonomen Lernenden Systems mit variablen Autonomiegraden



© Plattform Lernende Systeme

Als weitere wichtige Grundlagen für den Einsatz Lernender Systeme werden Autonomie, Steuerung und Lernen bei robotischen Systemen genannt, die beispielsweise grundlegende Steuerungsmodi für Roboter sowie Voraussetzungen für höhere Autonomiegrade robotischer Systeme betreffen. Entscheidende Komponenten von Autonomie bei Lernenden Systemen sind die Kenntnis der eigenen Fähigkeiten und die Bewertung der eigenen Handlung in Bezug zur Aufgabe, zur Umgebung und zur aktuellen Situation. Die größte Herausforderung für ein Lernendes System – welches vollkommen eigenständig agieren soll – besteht darin, all diese Parameter dynamisch über den Zeitraum des jeweiligen Einsatzes zu kennen und zusätzlich immer die Möglichkeit zu haben, auf neue Ereignisse zeitkritisch zu reagieren und das weitere Vorgehen darauf abzustimmen.

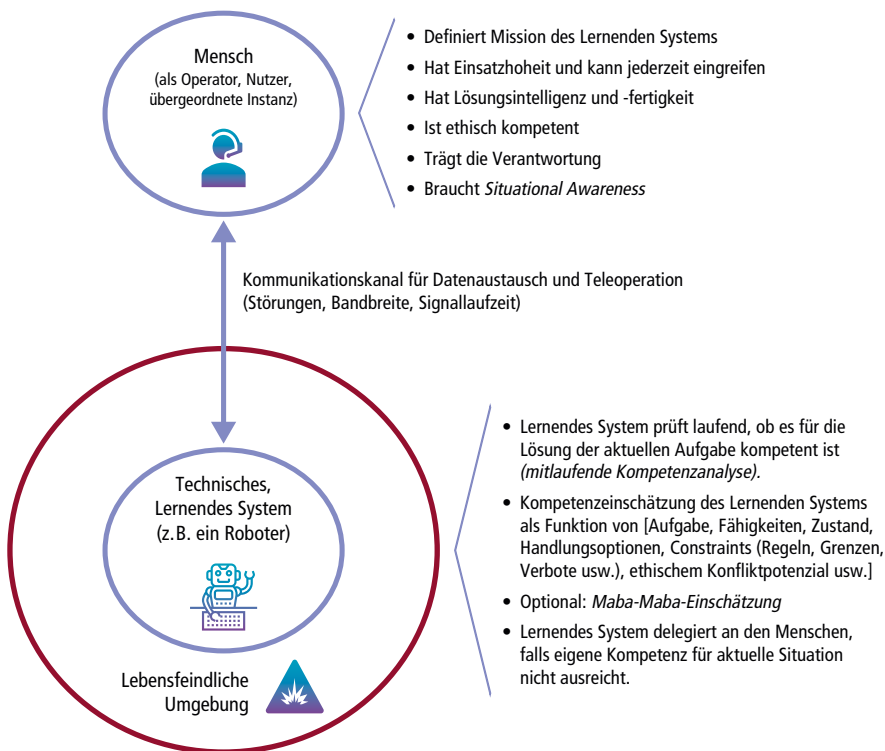
Variable Autonomiegrade von Lernenden Systemen

Ein Überblick über bereits existierende Stufenmodelle für Autonomiegrade Lernender Systeme in konventionellen Anwendungsdomänen wie Medizin oder autonomes Fahren macht deutlich, dass autonome Systeme in lebensfeindlichen Umgebungen mit besonderen Anforderungen konfrontiert sind, was zur Folge hat, dass auch die Anforderungen an Autonomie ganz andere sind. Die Einteilung bei existierenden Stufenmodellen für Autonomiegrade in konventionellen Bereichen reicht meist von vollständig (fern)gesteuert (nicht-autonom) bis hin zu vollständig autonom. Lebensfeindliche Umgebungen sind aber dagegen sehr vielfältig und Einsätze haben eine hohe Variabilität, verbunden mit sehr unterschiedlichen Anforderungen an die Autonomie in der jeweiligen Situation. Daher wird eine Erweiterung zur bisher bestehenden Unterscheidung diskreter Autonomiestufen hin zu variabler Autonomie skizziert. Um diese Perspektive der kontinuierlichen Autonomiegrade zur Laufzeit aufzuzeigen, konzipieren die Autorinnen und Autoren ein Architekturmodell, das Komponenten eines autonomen Systems definiert und verdeutlicht, wie verschiedene variable Autonomiegrade konkret aktiviert werden können.

Technische und ethische Vertrauenswürdigkeit Lernender Systeme

Grundsätzlich sollten Lernende Systeme mit Blick auf Kriterien – etwa Nachvollziehbarkeit, Sicherheit und der Vermeidung nicht-intendierter Folgewirkungen – ethisch vertrauenswürdig entwickelt werden. In lebensfeindlichen Umgebungen können Lernende Systeme während einer Mission auch auf ethisch oder rechtlich problematische Situationen treffen, die in der Folge auch die Kritikalität des Lernenden Systems in dessen Anwendungskontext in solchen Situationen steigern würde. Deshalb benötigen diese Systeme unter anderem Komponenten, die derartige Situationen erkennen und kommunizieren können. Die Kompetenz eines Lernenden Systems kann sich im Laufe des Einsatzes verändern, etwa weil es mit unterschiedlichen Anforderungen konfrontiert wird. Daher schlagen die Expertinnen und Experten eine Kompetenzanalyse zur Laufzeit für die Architektur Lernender Systeme vor, um sinnvolle und situationsbedingt mögliche Autonomiegrade bestimmen zu können. Beim Zusammenspiel von Kompetenzanalyse und Autonomiegrad bedeutet fehlende Kompetenz immer einen erhöhten Eingriff durch den Menschen bis hin zu einer Teleoperation (= Fernsteuerung). Bis zur vollständigen Autonomie lassen sich verschiedene Zwischenstufen unterscheiden, bei denen der Mensch nach und nach die Kontrolle an das Lernende System abgibt. Grundsätzlich kann der Fall eintreten, dass ein System eine Aufgabe beherrscht (z. B. Erkundung eines unbekanntes Terrains), es aber festgefahren ist und sich selbst nicht mehr befreien kann. In diesem Fall wären dessen Handlungsoptionen so weit eingeschränkt, dass eine Steuerung aus der Ferne durch den Menschen notwendig wird.

Lernende Systeme in lebensfeindlichen Umgebungen als Mensch-Maschine-Systeme



© Plattform Lernende Systeme

Gestaltungsoptionen und Ausblick

Damit Lernende Systeme den Menschen künftig noch effektiver bei Einsätzen in lebensfeindlichen Umgebungen unterstützen können, benennen die Expertinnen und Experten abschließend Gestaltungsoptionen, die sich vor allem auf die Stärkung von Forschung und Produktinnovationen beziehen. Forschungs- und Entwicklungsbedarf sehen sie vor allem hinsichtlich der Architekturkomponenten sowie der Kompetenzanalyse von Lernenden Systemen.

Architekturkomponenten

Als unverzichtbare Komponenten in der Architektur eines Mensch-Maschine-Systems für einen effektiven, effizienten und verantwortungsvollen Einsatz von Lernenden Systemen in lebensfeindlichen Umgebungen werden definiert:

- Mitlaufende Analyse, ob die Kompetenz des Lernenden Systems ausreicht, um eine aktuelle Aufgabe zu lösen.
- Mitlaufende Analyse, ob das Lernende System in eine ethisch problematische Situation gerät oder ethisch problematische Entscheidungen treffen müsste.
- Mitlaufende Analyse, ob das Lernende System in eine Situation gerät, die rechtliche Fragen aufwirft.
- Komponente für Sicherstellung von Situational Awareness für den Menschen als Voraussetzung für eine kurzfristige Kontrollübernahme.

Kompetenzanalyse

Um Lernende Systeme zu einer umfassenden Kompetenzanalyse als einer Art Selbst-Monitoring zu befähigen, sind mehrere Teilgebiete der Künstlichen Intelligenz (KI) zu integrieren. Etliche wichtige Meilensteine müssten hier erreicht werden, damit die Systeme diese Fähigkeit robust erlangen. Übergeordnet wären hierbei zu nennen:

- Das Lernende System schafft den Abgleich zwischen Aufgabe und eigenen Fähigkeiten: bei komplexen Aufgaben kann dies weitere Analysen mit einbeziehen, die heutige Systeme so noch nicht übergreifend leisten, wie etwa Analyse der Umgebung, Erfassung und Berücksichtigung von Kontexten, Generalisierung eigener Fähigkeit auf neue Aufgaben, Zerlegung einer komplexen Aufgabe in sinnvolle lösbare Teilaufgaben, effiziente Verteilung der Teilaufgaben unter mehreren Lernenden Systemen.
- Definition eines technischen Rahmens, wie Lernende Systeme...
 - ...ethisch problematische Situationen erkennen und kommunizieren können.
 - ...rechtlich problematische Vorgänge erkennen können.
- Sicherstellung von Situationsbewusstsein („Situational Awareness“) für den Menschen als Voraussetzung für eine kurzfristige Kontrollübernahme.

Letztlich betreffen all diese übergeordneten Meilensteine angelegte Forschungsfragen, auch die von Transparenz und Erklärbarkeit von KI, die wichtige Säulen der nationalen und europäischen KI-Strategien sind.

Für die Anwendenden gilt es, einen rechtlichen Rahmen zu schaffen, der Orientierung für Innovationen und Produkte geben kann. Unter welchen Bedingungen können Lernende Systeme für Einsätze in lebensfeindlichen Umgebungen genutzt werden? Was dürfen sie, was nicht? Diese Art von Vorgaben hilft auch den Bereichen der Kompetenzanalyse der Systeme selbst, zum Beispiel um festzustellen, wann eine Kompetenzanalyse durch das System selbst überhaupt notwendig wird. Nicht jedes System wird diese Fähigkeit gleichermaßen benötigen. Diese Fragestellungen geben dann auch wiederum Impulse in die Forschung. Letztlich stützt auch die Zertifizierung von Lernenden Systemen hier die Anwendung, da auch sie die Möglichkeit bietet, mehr Transparenz im Markt zu schaffen und Standards zu setzen.

Impressum

Herausgeber: Lernende Systeme – Die Plattform für Künstliche Intelligenz | Geschäftsstelle | c/o acatech | Karolinenplatz 4 | D-80333 München | kontakt@plattform-lernende-systeme.de | www.plattform-lernende-systeme.de | Folgen Sie uns auf Twitter: @LernendeSysteme | Stand: Januar 2021 | Bildnachweis: designprojects/AdobeStock/Titel

Diese Kurzfassung entstand auf Grundlage des Whitepapers *Kompetent im Einsatz – Variable Autonomie Lernender Systeme in lebensfeindlichen Umgebungen*, München, 2021. Es wurde erstellt von einem Autorenteam der Arbeitsgruppe Lebensfeindliche Umgebungen der Plattform Lernende Systeme. Die Originalfassung der Publikation ist online verfügbar unter: <https://www.plattform-lernende-systeme.de/publikationen.html>



GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

 acatech
DEUTSCHE AKADEMIE DER
TECHNIKWISSENSCHAFTEN